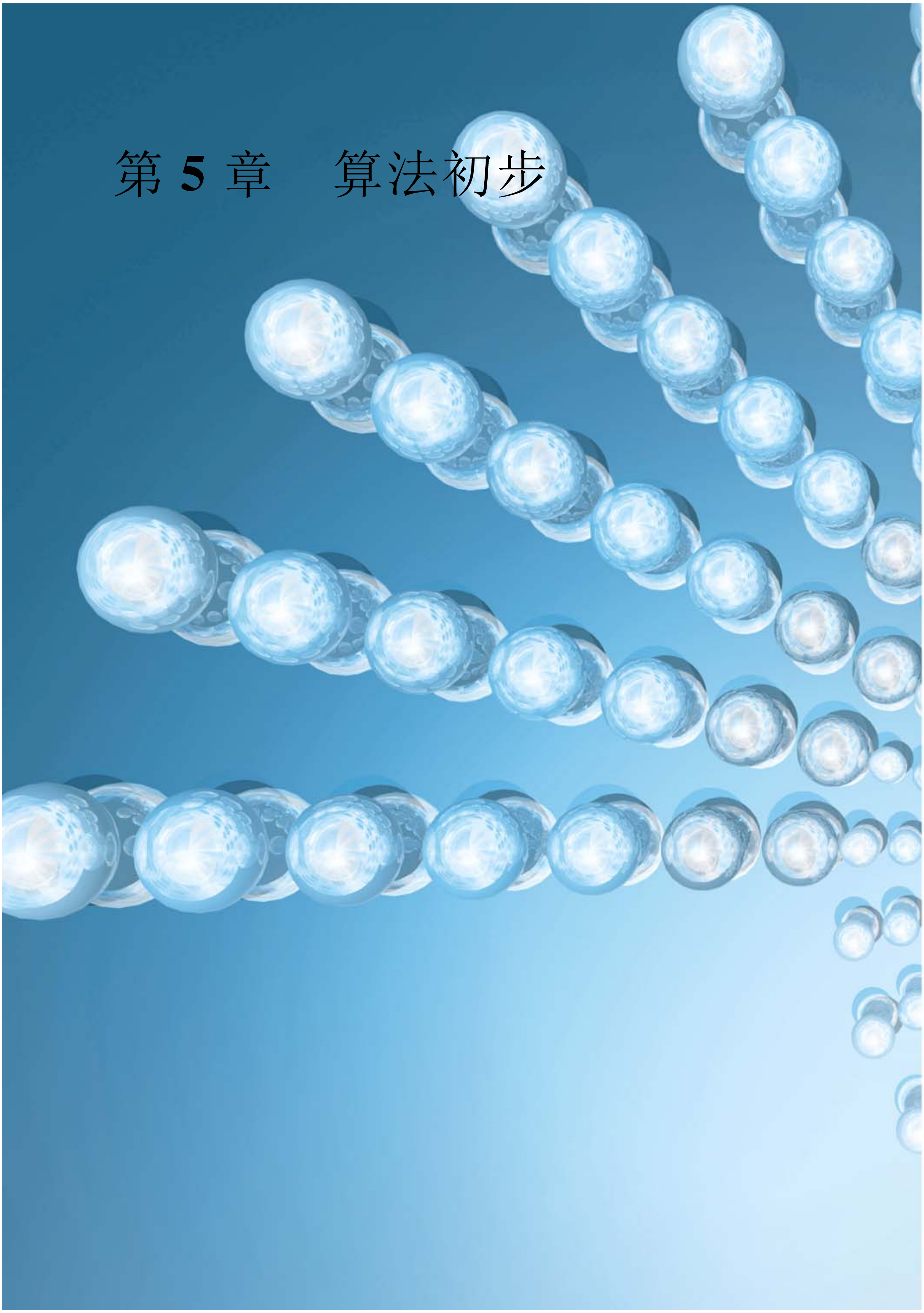
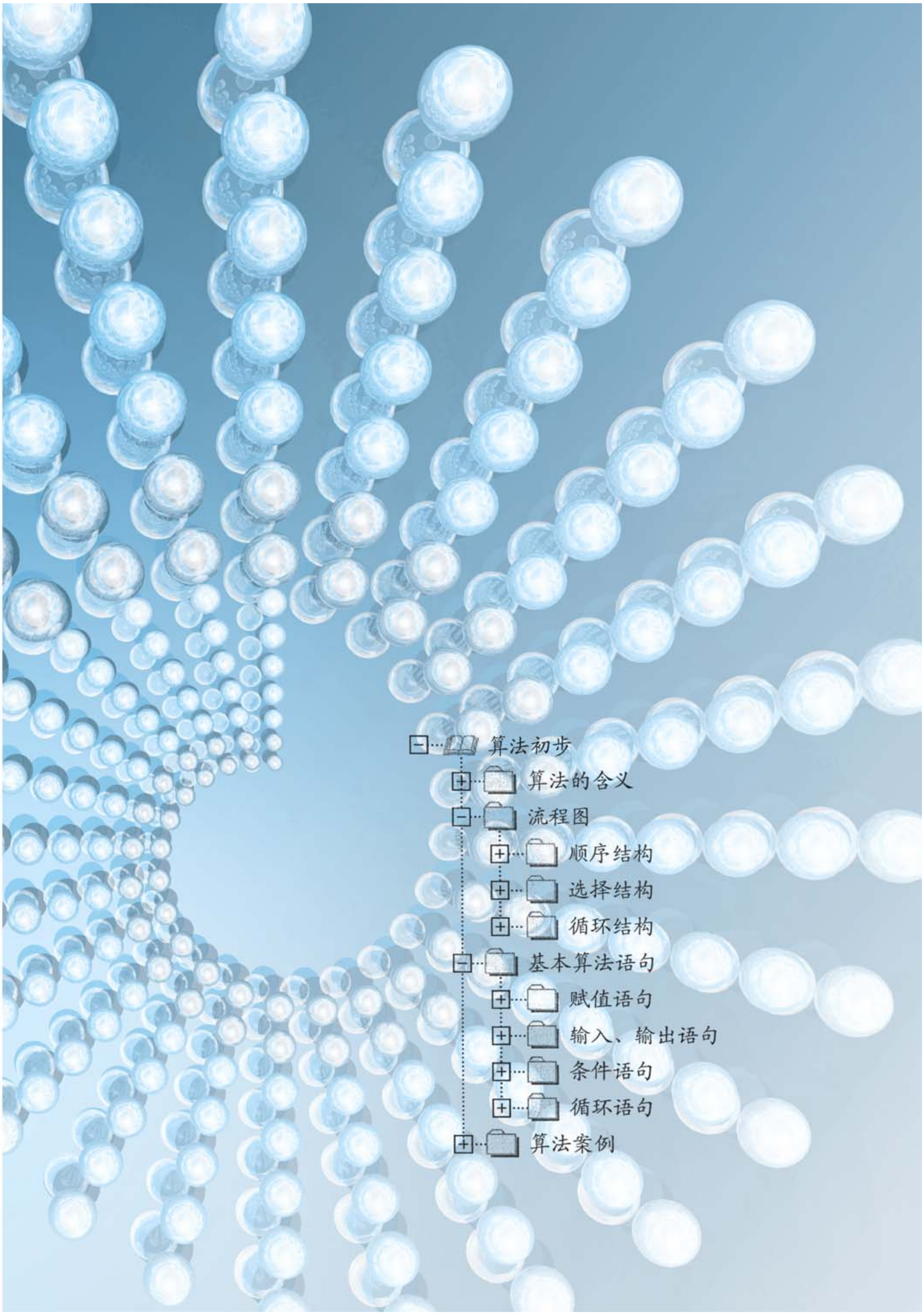
目　　录

第 5 章　算法初步

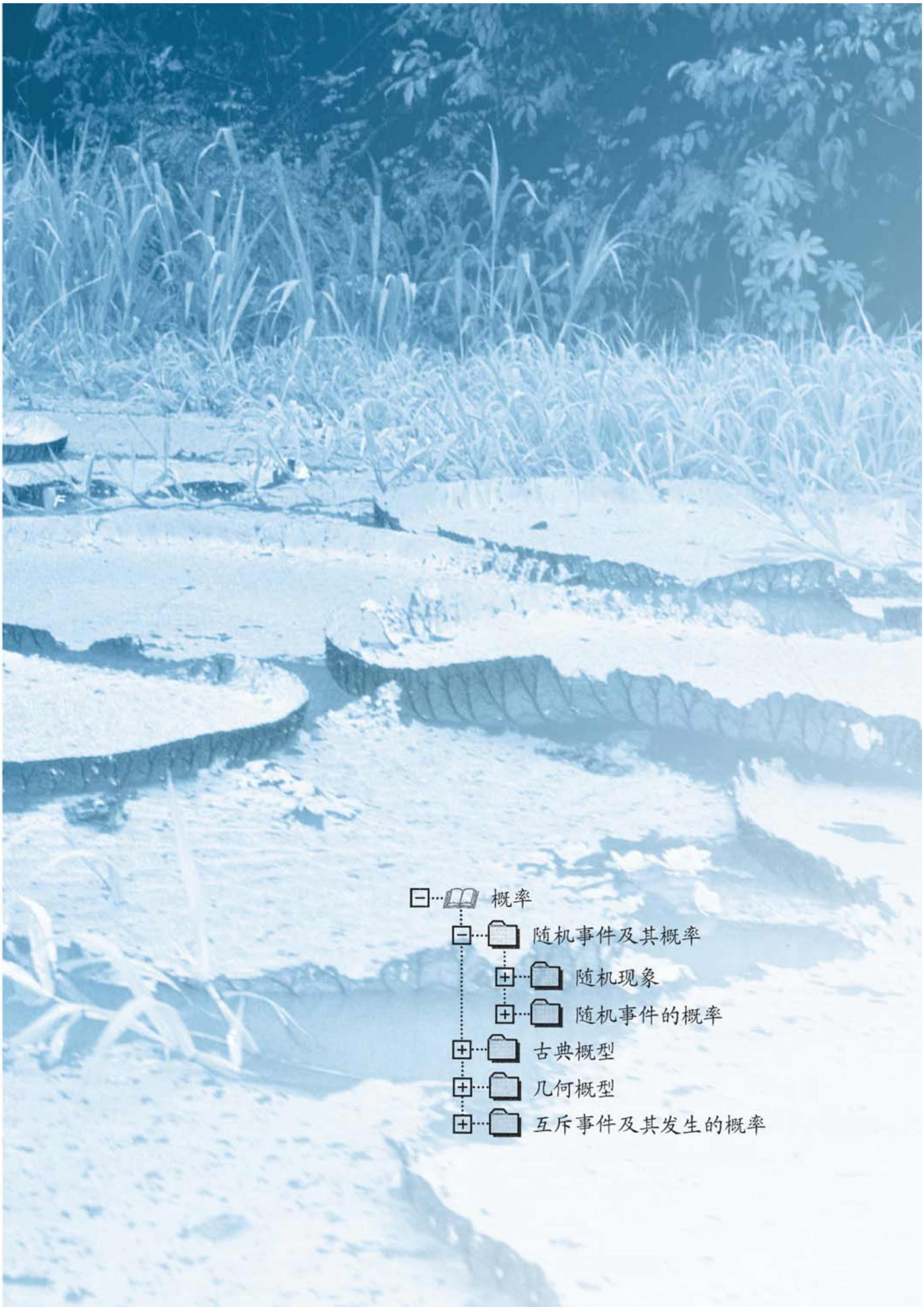



算法初步
算法的含义
流程图
顺序结构
选择结构
循环结构
基本算法语句
赋值语句
输入、输出语句
条件语句
循环语句
算法案例

第 6 章　统　　计

- 统计
  - 抽样方法
    - 简单随机抽样
      - 抽签法
      - 随机数表法
    - 系统抽样
    - 分层抽样
  - 总体分布的估计
    - 频率分布表
    - 频率分布直方图与折线图
    - 茎叶图
  - 总体特征数的估计
    - 平均数及其估计
    - 方差与标准差
  - 线性回归方程

第 7 章　概　　率



- 概率
  - 随机事件及其概率
    - 随机现象
    - 随机事件的概率
  - 古典概型
  - 几何概型
  - 互斥事件及其发生的概率

数学是科学的大门和钥匙.

——伽利略

一种科学只有在成功地运用数学时，才算达到完善的地步.

——马克思

致　同　学

亲爱的同学，你感到高中阶段的学习生活有趣吗？

我们知道，数学与生活紧密相连. 数学可以帮助我们认识世界，改造世界，创造新的生活. 数学是高中阶段的重要学科，不仅是学习物理、化学等学科的基础，而且对我们的终身发展有较大的影响.

面对实际问题，我们要认真观察、实验、归纳，大胆提出猜想. 为了证实或推翻提出的猜想，我们要通过分析，概括、抽象出数学概念，通过探究、推理，建立数学理论. 我们要积极地运用这些理论去解决问题. 在探究与应用过程中，我们的思维水平会不断提高，我们的创造能力会得到发展. 在数学学习过程中，我们将快乐地成长.

考虑广大同学的不同需要，本书提供了较大的选择空间.

书中的引言、正文、练习、习题中的“感受・理解”部分、阅读、回顾等内容构成一个完整的体系. 它体现了教材的基本要求，是所有学生应当掌握的内容. 相信你一定能学好这部分内容.

本书还设计了一些具有挑战性的内容，包括思考、探究、链接，以及习题中的“思考・运用”、“探究・拓展”等，以激发你探索数学的兴趣. 在掌握基本内容之后，选择其中一些内容作思考与探究，你会更加喜欢数学.

本书部分常用符号

	输入、输出框，如输入 n
	处理框，如计算 n^2 的值
	判断框，如判断是否大于 2 004
	起止框
$p \leftarrow q$	把 q 赋给 p
$\overline{x}$	样本平均数
$\sum\limits_{i=1}^{n} a_i$	n 个实数 a_1，a_2，…，a_n 的和
$P(A)$	事件 A 的概率
$\overline{A}$	事件 A 的对立事件

对世界各式各样的观察方式中，……最有趣的方式之一，是被设想为由模式组成的那种方式．

——维　纳

假如你的朋友不会发电子邮件，你能教会他吗？发邮件的方法很多，下面就是一种操作步骤：

第一步　打开电子信箱；

第二步　点击“写邮件”；

第三步　输入发送地址；

第四步　输入主题；

第五步　输入信件内容；

第六步　点击“发送邮件”.

我们做任何一件事，都是在一定的条件下按某种顺序执行的一系列操作. 解决数学问题也常常如此. 例如，用加减消元法解二元一次方程组时，就可以按照某一程序进行操作；用配方法解一元二次方程，也是按一定程序操作的.

将上述程序转换成计算机能识别的语言后，就能借助计算机极大地提高解决问题的速度. 因此，探索解决问题的统一程序的思想是十分重要的. 对一类问题的机械的、统一的求解程序就是算法.

从古代的“百鸡问题”到现代机器证明数学定理的“吴方法”，从二元一次方程组的消元解法到计算机动画的设计，从猜数游戏到集成电路的布线安排，……它们蕴含了丰富的算法思想.

面对一个需要解决的问题，

- 如何设计解决问题的操作序列？
- 怎样用数学语言描述这些操作序列？

5.1 算法的含义

电视娱乐节目中，有一种有趣的“猜数”游戏：竞猜者如在规定的时间内猜出某种商品的价格(或重量等)，就可获得该件商品.

● 现有一商品，价格在 0～8 000 元之间，采取怎样的策略才能在较短的时间内说出正确的答案呢？

解决这个问题有多种途径，其中一种较好的方法是：

第一步　报“4 000”；

第二步　若主持人说“高了”(说明答数在 1～4 000 之间)，就报“2 000”，否则(答数在 4 000 到 8 000 之间)报“6 000”；

第三步　重复第二步的报数方法，直至得到正确结果.

以上过程实际上是按一种机械的程序进行的一系列的操作.

一般而言，对一类问题的机械的、统一的求解方法称为**算法**(algorithm).

阿尔·花拉子米(al-Khowârizmî，约780～约850)，主要著作有《代数学》，被西方人称为“代数学之父”.英文 algorithm 一词据说是源于这位数学家的名字.

例 1　给出求 $1+2+3+4+5$ 的一个算法.

算法 1　按照逐一相加的程序进行.

第一步　计算 $1+2$，得到 3；

第二步　将第一步中的运算结果 3 与 3 相加，得到 6；

第三步　将第二步中的运算结果 6 与 4 相加，得到 10；

第四步　将第三步中的运算结果 10 与 5 相加，得到 15.

算法 2　可以运用公式 $1+2+3+\cdots+n=\dfrac{n(n+1)}{2}$ 直接计算.

第一步　取 $n=5$；

第二步　计算 $\dfrac{n(n+1)}{2}$；

第三步　输出运算结果.

例 2　给出求解方程组

$$\begin{cases}2x+y=7, & ① \\ 4x+5y=11 & ②\end{cases}$$

的一个算法.

解　我们用消元法求解这个方程组，步骤是：

第一步　方程①不动，将方程②中 x 的系数除以方程①中 x 的

系数，得到乘数

$$m=\frac{4}{2}=2;$$

第二步　方程②减去 m 乘以方程①，消去方程②中的 x 项，得到

$$\begin{cases}2x+y=7,\\ \quad 3y=-3;\end{cases}$$

第三步　将上面的方程组自下而上回代求解，得到

$$y=-1,\ x=4.$$

所以原方程组的解为

$$\begin{cases}x=4,\\ y=-1.\end{cases}$$

这种消元回代的算法适用于一般线性方程组的求解.

找到了某种算法，是指使用一系列运算规则能在有限步骤内求解某类问题，其中的每条规则必须是明确定义的、可行的.

算法从初始步骤开始，每一个步骤只能有一个确定的后继步骤，从而组成一个步骤序列，序列的终止表示问题得到解答或指出问题没有解答.

我们过去学过的许多数学公式都是算法，加、减、乘、除运算法则以及多项式的运算法则也是算法.

练　习

1. 写出解方程 $2x+3=0$ 的一个算法过程.
2. 写出求 $1\times3\times5\times7$ 的一个算法.
3. 已知直角坐标系中的两点 $A(-1,0)$，$B(3,2)$，写出求直线 AB 的方程的一个算法.
4. 写出求 $1+2+3+\cdots+100$ 的一个算法.

5.2 流程图

● 回答下面的问题：

(1) $1+2+3+\cdots+100=$ ________；

(2) $1+2+3+\cdots+n=$ ________；

(3) $1+2+3+\cdots+$ ________ $>2\,004$.

第(3)个问题的答案并不惟一，为了寻找满足条件的最小正整数，我们可以这样设计算法：

这里 S1 代表步骤 1，S2 代表步骤 2，以此类推. S 是 step（步）的第一个字母.

S1　取 n 等于 1；

S2　计算 $\dfrac{n(n+1)}{2}$；

S3　如果 $\dfrac{n(n+1)}{2}$ 的值大于 2 004，那么 n 即为所求；否则让 n 的值增加 1 后转到 S2 重复操作.

思　考

如果在上面的 S1 中，给出的 n 的初始值不是 1，而是一个较大的值（如 $n=2\,000$），为了求出满足条件的最小正整数，应该如何设计算法？

将设计好的算法清晰直观地描述出来，通常采用图 5－2－1 画流程图的形式来表示.

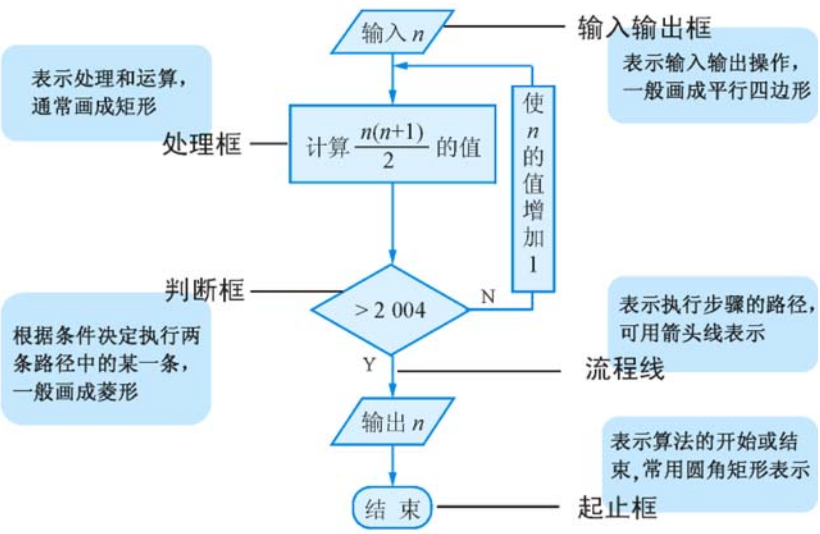


图 5－2－1

流程图(flow chart)是由一些图框和带箭头的流线组成的，其中图框表示各种操作的类型，图框中的文字和符号表示操作的内容，带箭头的流线表示操作的先后次序.

从流程图 5－2－1 中可以看出，该算法步骤中，有些是按顺序执行，有些需要选择执行，而另外一些需要循环执行.

事实上，算法都可以由顺序结构、选择结构、循环结构这三块“积木”通过组合和嵌套表达出来. 流程图可以帮助我们更方便直观地表示这三种基本的算法结构.

5.2.1 顺序结构

● 写出作△ABC 的外接圆的一个算法.

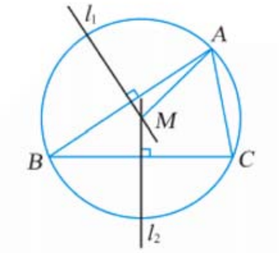


图 5－2－2

解决这个问题可按下面的算法进行：

S1 作 AB 的垂直平分线 l_1；

S2 作 BC 的垂直平分线 l_2；

S3 以 l_1 与 l_2 的交点 M 为圆心，MA 为半径作圆，圆 M 即为△ABC 的外接圆.

以上过程通过依次执行 S1 到 S3 这三个步骤，完成了作外接圆这一问题(图 5－2－2). 像这种依次进行多个处理的结构称为顺序结构(sequence structure). 如图 5－2－3 所示，虚线框内是一个顺序结构，其中 A 和 B 两个框是依次执行的. 顺序结构是一种最简单、最基本的结构.

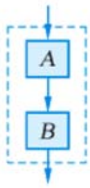


图 5－2－3

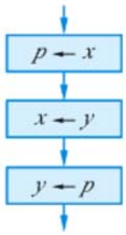


图 5－2－4

例 1 已知两个单元分别存放了变量 x 和 y 的值，试交换这两个变量值.

解 为了达到交换的目的，需要一个单元存放中间变量 p. 其算法是：

在计算机中，每个变量都分配了一个存储单元，它们都有各自的“门牌号码”(地址).

S1 $p \leftarrow x$； {先将 x 的值赋给变量 p，这时存放变量 x 的单元可作它用}

S2 $x \leftarrow y$； {再将 y 的值赋给 x，这时存放变量 y 的单元可作它用}

S3 $y \leftarrow p$. {最后将 p 的值赋给 y，两个变量 x 和 y 的值便完成了交换}

上述算法用流程图表示如图 5－2－4.

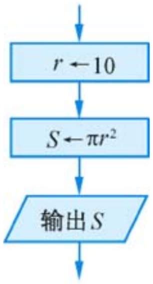


图 5-2-5

例 2 半径为 r 的圆的面积计算公式为

$$S=\pi r^2,$$

当 $r=10$ 时，写出计算圆面积的算法，画出流程图.

解 算法如下：

S1 $r \leftarrow 10$； {把 10 赋给变量 r}

S2 $S \leftarrow \pi r^2$； {用公式计算圆的面积}

S3 输出 S. {输出圆的面积}

上述算法用流程图表示如图 5-2-5.

练 习

1. 写出作棱长为 2 的正三棱柱的直观图的算法.

2. 写出解方程组 $\begin{cases} x+y=3, \\ y+z=5, \\ z+x=4 \end{cases}$ 的一个算法，并用流程图表示算法过程.

5.2.2 选择结构

● 某铁路客运部门规定甲、乙两地之间旅客托运行李的费用为

$$c=\begin{cases} 0.53\times w, & w\leqslant 50, \\ 50\times 0.53+(w-50)\times 0.85, & w>50, \end{cases}$$

其中 w（单位：kg）为行李的重量，试给出计算费用 c（单位：元）的一个算法.

为了计算行李的托运费，应先判断行李的重量是否大于 50 kg，然后再选用相应的公式进行计算. 其算法为：

S1 输入行李的重量 w；

S2 如果 $w\leqslant 50$，那么 $c=0.53\times w$，

否则 $c=50\times 0.53+(w-50)\times 0.85$；

S3 输出行李重量 w 和运费 c.

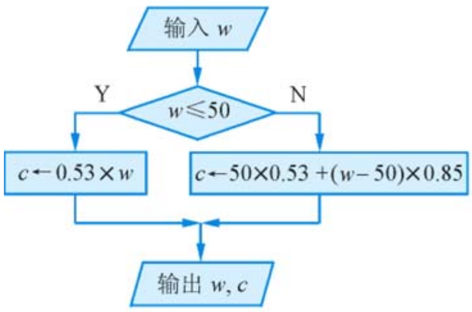


图 5-2-6

在上述计费过程中，第二步我们进行了判断，像这种先根据条件作出判断，再决定执行哪一种操作的结构称为选择结构（selection

structure)(或称为“分支结构”). 如图 5－2－7 所示,虚线框内是一个选择结构,它包含一个判断框,当条件 p 成立(或称为“真”)时执行 A,否则执行 B.

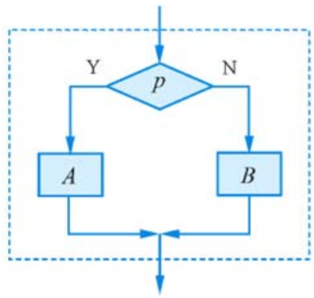


图 5－2－7

思　考

在流程图 5－2－1 所表示的算法中,哪一步进行了判断?

例 3　设计求解一元二次方程

$$ax^2+bx+c=0\ (a\neq 0)$$

的一个算法,并用流程图表示.

分析　由于一元二次方程未必总有实数根,因此求解时,要先计算判别式 $\Delta=b^2-4ac$,然后比较判别式与 0 的大小,再决定能否用求根公式进行求解. 因此,在算法中应含有选择结构.

解　算法如下:

S1　输入 a, b, c;

S2　$\Delta \leftarrow b^2-4ac$;

S3　如果 $\Delta<0$ 为真,则输出“方程无实数根”,否则

$$x_1\leftarrow\frac{-b+\sqrt{\Delta}}{2a},\ x_2\leftarrow\frac{-b-\sqrt{\Delta}}{2a}.$$

根据上述步骤,我们可以画出如图 5－2－8 所示的算法流程图.

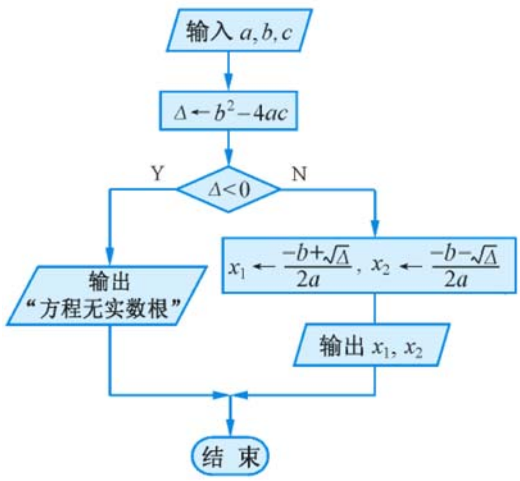
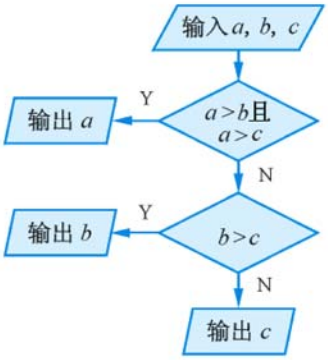


图 5－2－8

练 习

1. 如果考生的成绩大于或等于 60 分，则输出“及格”，否则输出“不及格”. 用流程图表示这一算法过程.
2. 下面的流程图表示了一个什么样的算法？



（第 2 题）

3. 写出解方程 $ax+b=0$（a，b 为常数）的算法，并画出流程图.

5.2.3 循环结构

北京获得了 2008 年第 29 届奥林匹克运动会主办权. 你知道在申办奥运会的最后阶段，国际奥委会是如何通过投票决定主办权归属的吗？

对遴选出的 5 个申办城市进行表决的操作程序是：首先进行第一轮投票，如果有一个城市得票超过总票数的一半，那么该城市将获得举办权；如果所有申办城市得票数都不超过总票数的一半，则将得票数最少的城市淘汰，然后重复上述过程，直到选出一个申办城市为止.

● 用怎样的算法结构表述上面的操作过程？

我们分步描述操作过程：

S1 投票；

S2 统计票数，如果有一个城市得票数超过总票数的一半，那么该城市就获得主办权，转 S3，否则淘汰得票数最少的城市，转 S1；

S3 宣布主办城市.

上述过程可用图 5－2－9 所示的流程图来表示.

在算法中，像这种需要重复执行同一操作的结构称为循环结构(cycle structure). 图 5－2－10 就是常见的一种循环结构：先执行 A 框，再判断给定的条件 p 是否为“假”；若 p 为“假”，则再执行 A，如此反复，直到 p 为“真”，该循环过程结束.

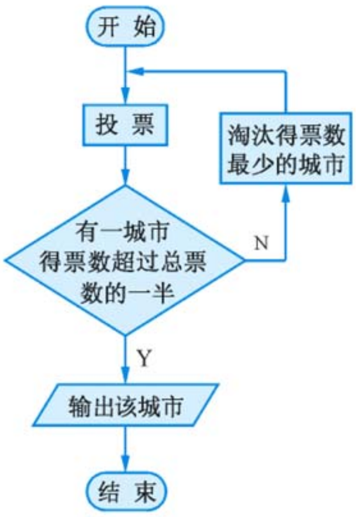


图 5－2－9

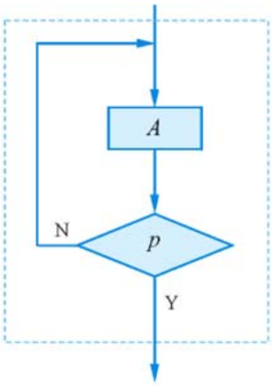


图 5－2－10

思　考

在流程图 5－2－1 中，哪些步骤构成了循环结构？

例 4　写出求 $1\times2\times3\times4\times5$ 值的一个算法.

算法 1

S1　先求 1×2，得到 2；

S2　将 S1 得到的结果再乘以 3，得到 6；

S3　将 S2 得到的结果 6 再乘以 4，得到 24；

S4　将 S3 得到的结果 24 再乘以 5，得到最后的结果 120.

上述算法虽然正确，但在计算 $1\times2\times\cdots\times100$ 时，算法的程序太长.

算法 2

S1　$T\leftarrow1$；　　{使 $T=1$}

S2　$I\leftarrow2$；　　{使 $I=2$}

S3　$T\leftarrow T\times I$；　　{求 $T\times I$，乘积结果仍放在变量 T 中}

S4　$I\leftarrow I+1$；　　{使 I 的值增加 1}

S5　如果 I 不大于 5，返回重新执行步骤 S3 及 S4，S5，否则算法结束. 这样最后得到的 T 值就是所要求的结果.

对算法 2 做少许改动，求 $1\times3\times5\times7\times9\times11$ 的值.

算法 2 不仅形式简练，而且具有通用性、灵活性. 其中，S3 到 S5 组成一个循环，在实现算法时要反复多次执行 S3，S4，S5 步骤，直到执行 S5 步骤时，经过判断，乘数 I 已超过规定的数为止. 其流程图如图 5－2－11 所示.

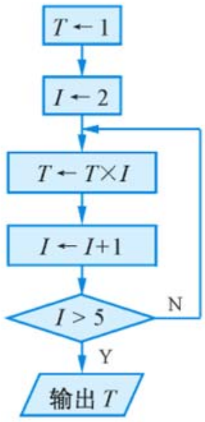


图 5－2－11

例 5 设计一个计算 10 个数平均数的算法.

分析 我们用一个循环依次输入 10 个数,再用一个变量存放数的累加和,在求出 10 个数的总和后,除以 10,就得到 10 个数的平均数.

将 0 赋值于 S,是为这些数的和建立存放空间.

解 S1 $S \leftarrow 0$; {使 $S=0$}

S2 $I \leftarrow 1$; {使 $I=1$}

S3 输入 G; {输入一个数}

S4 $S \leftarrow S+G$; {求 $S+G$,其和仍放在 S 中}

S5 $I \leftarrow I+1$; {使 I 的值增加 1 后循环}

S6 如果 I 不大于 10,转 S3; {如果 $I>10$,退出循环}

S7 $A \leftarrow S/10$; {将平均数 $S/10$ 存放到 A 中}

S8 输出 A. {输出平均数}

算法流程图如图 5－2－12.

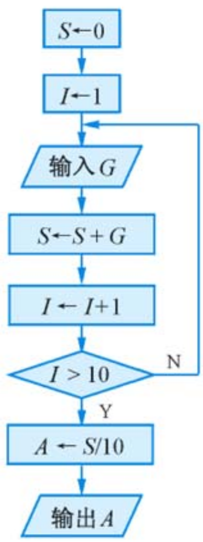
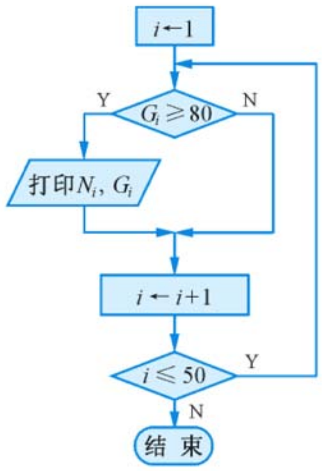


图 5－2－12

由上面的例题可以看到，利用顺序结构、选择结构和循环结构这三种基本结构描述的算法，结构清晰，容易阅读、理解和修改.

练 习

1. 先分步写出计算 $2+4+6+\cdots+100$ 的一个算法，再画出流程图.
2. 用 N_i 代表第 i 个学生的学号，G_i 代表第 i 个学生的成绩 $(i=1, 2, \cdots, 50)$，那么下图表示了一个什么样的算法？



（第 2 题）

习题 5.1

感受 · 理解

1. 三角形面积的计算公式为 $S=\frac{1}{2}ah$（其中 a 为边长，h 为该边上的高），用算法描述求 $a=7.85$，$h=14.29$ 时的三角形面积，并画出流程图.
2. 火车站对乘客在一定时段内退票要收取一定的费用，收费的办法是：按票价每 10 元(不足 10 元按 10 元计算)核收 2 元，2 元以下的票价不退. 试分步写出将票价为 x 元的车票退掉后，返还的金额 y 的算法，并画出流程图.
3. 画出解方程组 $\begin{cases}2x-y=1,\\4x+3y=7\end{cases}$ 的一个算法的流程图.
4. 画出求两个正整数 a 与 b 相除所得商 q 及余数 r 的一个算法的流程图.
5. 写出一个求三个实数中最小数的算法，并用流程图表示.
6. 写出解不等式 $ax+b>0$ $(a\neq 0)$ 的一个算法，并画出流程图.
7. 写出求 $\cfrac{1}{2+\cfrac{1}{2+\ddots+\cfrac{1}{2}}}$（共有 6 个 2）的值的一个算法，并画出流程图.

思考 · 运用

8. 写出在数 3，5，8，9，12，15，35，7，18，52 中搜索数 18 的一个算法，并画出流程图.

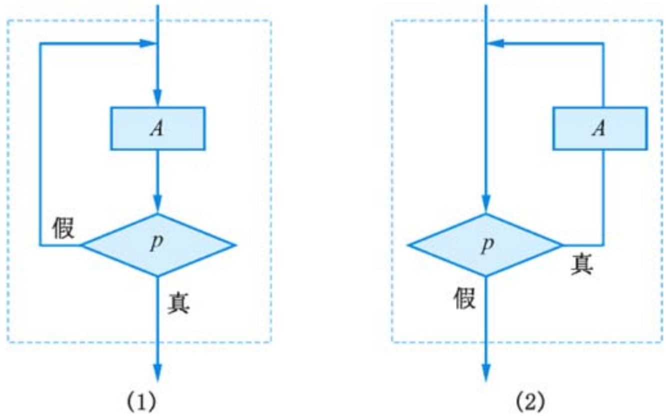
海伦（Heron，生卒年不详），古希腊数学家，约公元 62 年前后活跃于亚历山大.

9. 我国南宋时期的数学家秦九韶发现了求三角形面积的“三斜求积”公式 $S_{\triangle ABC}=\sqrt{\frac{1}{4}\left[c^2a^2-\left(\frac{c^2+a^2-b^2}{2}\right)^2\right]}$，它与古希腊数学家海伦给出的三角形面积公式 $S_{\triangle ABC}=\sqrt{p(p-a)(p-b)(p-c)}$ $\left(其中\ p=\frac{1}{2}(a+b+c)\right)$ 是一致的. 试根据海伦公式画出已知三边 a，b，c，求三角形面积的流程图.（提示：输入 3 个数后应先判断它们能否构成三角形的三边）

探究・拓展

10.（阅读题）常见的循环结构有两种：一种是前面我们介绍的直到型循环结构，另一种是当型循环结构.

直到型（Until 型）循环：如图（1），先执行 A 框，再判断给定的条件 p 是否为“假”，若 p 为“假”，则再执行 A，如此反复，直到 p 为“真”为止.



（第 10 题）

当型（While 型）循环：如图（2），当给定的条件 p 成立（“真”）时，反复执行 A 框操作，直到条件 p 为“假”时才停止循环.

当型循环和直到型循环是可以互相转换的. 试用当型循环写出例 4 的算法步骤，并画出流程图.

5.3 基本算法语句

● 算法是一种数学语言，如何用更简捷的语句表述算法语言呢？

本节主要通过伪代码学习基本的算法语句.

伪代码(pseudo code)是介于自然语言和计算机语言之间的文字和符号，是表达算法的简单而实用的好方法. 下面，我们在伪代码中将使用 BASIC 语言的关键词.

5.3.1 赋值语句

在伪代码中，赋值语句(assignment statement)用符号“$\leftarrow$”表示，“$x \leftarrow y$”表示将 y 的值赋给 x，其中 x 是一个变量，y 是一个与 x 同类型的变量或表达式.

例 1 写出求 $x=23$ 时多项式 $7x^3+3x^2-5x+11$ 的值的算法.

算法 1

$x \leftarrow 23$;

$p \leftarrow 7x^3+3x^2-5x+11$.

算法 2

$x \leftarrow 23$;

$p \leftarrow ((7x+3)x-5)x+11$.

上述两种算法，算法 1 要做 6 次乘法，算法 2 只需做 3 次乘法. 由此可见，算法的好坏会影响运算速度.

例 1 中算法 2 称为秦九韶算法，其算法特点是：通过一次式的反复计算，逐步得出高次多项式的值；对于一个 n 次多项式，只要做 n 次乘法和 n 次加法.

5.3.2 输入、输出语句

● “鸡兔同笼”是我国隋朝时期的数学著作《孙子算经》中的一个有趣而具有深远影响的题目：

“今有雉兔同笼，上有三十五头，下有九十四足. 问雉兔各几何.”

用方程组的思想不难解决这一问题.

设有 x 只鸡，y 只兔，则

$$\begin{cases} x+y=35, \\ 2x+4y=94. \end{cases}$$

下面我们设计一个解二元一次方程组的通用算法.

设二元一次方程组为

$$\begin{cases} a_1x+b_1y=c_1, \\ a_2x+b_2y=c_2, \end{cases} (a_1b_2-a_2b_1\neq 0)$$

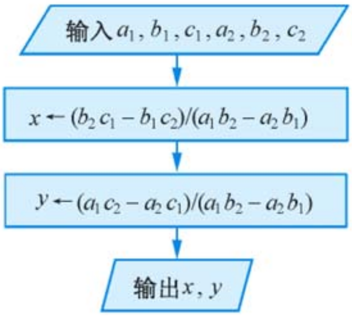
《孙子算经》(局部)

用消元法解得

$$\begin{cases} x=\dfrac{b_2c_1-b_1c_2}{a_1b_2-a_2b_1}, \\ y=\dfrac{a_1c_2-a_2c_1}{a_1b_2-a_2b_1}, \end{cases}$$

因此,只要输入相应的未知数的系数和常数项,就能计算出方程组的解,即可以输出 x, y 的值.

我们用输入语句(input statement)“Read　a, b”表示输入的数据依次送给 a, b,用输出语句(output statement)“Print　x”表示输出运算结果 x. 这样上述解二元一次方程组的算法的流程图与相应的伪代码就可以表示为图 5-3-1.




```
Read  a1, b1, c1, a2, b2, c2
x ← (b2c1−b1c2)/(a1b2−a2b1)
y ← (a1c2−a2c1)/(a1b2−a2b1)
Print  x, y
```

图 5-3-1

当输入 a_1, b_1, c_1, a_2, b_2, c_2 分别为 1, 1, 35, 2, 4, 94 时,输出的 x, y 的值分别为 23, 12. 即“鸡兔同笼”问题的答案是 23 只鸡和 12 只兔.

练　习

1. 已知一个正三棱柱的底面边长为 2,高为 3,用输入、输出语句和赋值语句表示计算此三棱柱的体积的算法.

2. 若三角形的三边长分别为 a, b, c,借助三角形的面积公式

$$S=\sqrt{p(p-a)(p-b)(p-c)}\ \left(\text{其中 } p=\frac{1}{2}(a+b+c)\right),$$

用输入、输出语句和赋值语句表示计算三角形面积的一种算法.

3. 某市 2003 年 1～12 月的产值分别为 3.8, 4.2, 5.3, 6.1, 5.6, 4.8, 7.3, 4.5, 6.4, 5.8, 4.7, 6.5(亿元),该市要统计每季度的月平均产值及 2003 年的月平均产值,试分别用赋值语句和输入、输出语句表示计算上述各个平均值的算法.

5.3.3 条件语句

某居民区的物管部门每月按以下方法收取卫生费：3 人和 3 人以下的住户，每户收取 5 元；超过 3 人的住户，每超出 1 人加收 1.2 元.

● 如何设计算法，根据输入的人数计算应收取的卫生费？

我们令 c（单位：元）表示应收取的费用，t 表示这户人家的人口数，则有

$$c=\begin{cases}5, & 0<t\leqslant 3,\\ 5+1.2(t-3), & t>3.\end{cases}$$

解决这一问题的算法步骤如下：

S1 输入人数 t；

S2 如果 $t\leqslant 3$，那么 $c\leftarrow 5$，
否则 $c\leftarrow 5+1.2(t-3)$；

S3 输出 c.

其流程图如图 5 - 3 - 2 所示.

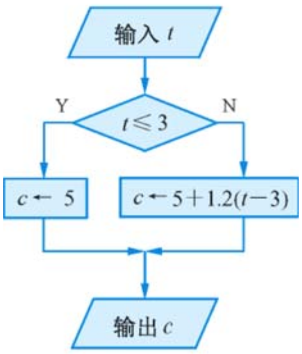


图 5 - 3 - 2

从流程图 5 - 3 - 2 可以看出这是一个选择结构. 在执行此算法时，要根据一定的条件选择箭头流线的方向.

我们可运用条件语句(conditional statement)来实现上述过程. 条件语句的一般形式是

```
If A then B
Else C
End if
```

其中 A 表示判断的条件，B 表示满足条件时执行的操作内容，C 表示不满足条件时执行的操作内容，End if 表示条件语句结束.

上面的算法过程用条件语句可表示为

```
Read t
If t≤3 then
    c ← 5
Else
    c ← 5+1.2(t−3)
End if
Print c
```

我们把步骤"$c\leftarrow 5$"称为"then"分支，步骤"$c\leftarrow 5+1.2(t-3)$"称为"Else"分支. 为了醒目和便于阅读，这些分支一般缩进书写.

例 2 儿童乘坐火车时，若身高不超过 1.1 m，则无需购票；若身高超过 1.1 m 但不超过 1.4 m，可买半票；若超过 1.4 m，应买全票. 试设计一个购票的算法，写出伪代码，并画出流程图.

解 上述购票的算法步骤为：

S1 测量儿童身高 h；

S2 如果 $h\leqslant 1.1$，那么免费乘车；否则，如果 $h\leqslant 1.4$，那么购半票乘车；否则，购买全票.

用条件语句表示为：

```
Read h
If h≤1.1 then
   Print 免费乘车
Else if h≤1.4 then
   Print 半票乘车
Else
   Print 全票乘车
End if
```

流程图如图 5－3－3 所示.

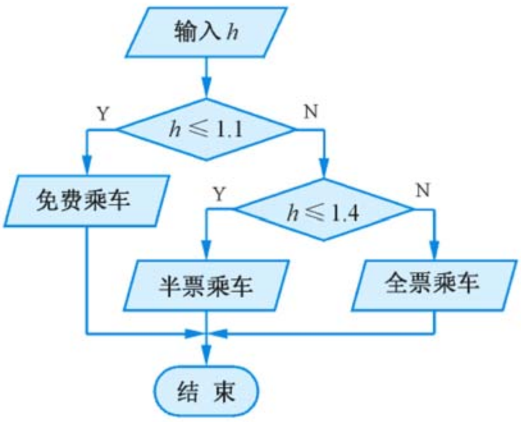


图 5－3－3

由例 2 可知，条件语句“If-then-Else” 可以嵌套.

思 考

条件语句也可以没有“Else”分支. 你能举一个例子说明吗？

例 3 已知函数

$$y=\begin{cases}1, & x>0,\\ 0, & x=0,\\ -1, & x<0,\end{cases}$$

试写出计算 y 值的算法.

解　用条件语句可以方便地表示这类分段函数：

```
Read x
If x > 0 then
  y ← 1
Else if x = 0 then
  y ← 0
Else
  y ← −1
End if
Print y
```

对应的流程图如图 5-3-4 所示.

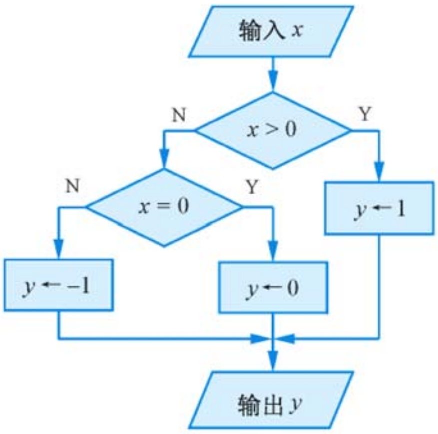


图 5-3-4

本例中的分段函数称为“符号函数”，即通过函数值(输出值)是 1 还是 −1 来判断输入的值是正数还是负数.

练　习

1. 用条件语句表示：输入两个数，打印较大的数.

2. 已知函数 $y=\begin{cases} x, & x\geqslant 0, \\ -x, & x<0, \end{cases}$ 试写出计算 y 值的算法.

3. 到银行办理个人异地汇款(不超过 100 万)时，银行要收取一定的手续费. 汇款额不超过 100 元，收取 1 元手续费；超过 100 元但不超过 5 000 元，按汇款额的 1%收取；超过 5 000 元，一律收取 50 元手续费. 试用条件语句描述汇款额为 x(元)时，求银行收取的手续费 y(元)的算法过程，并画出流程图.

5.3.4 循环语句

● 设计计算

$$1\times3\times5\times7\times\cdots\times99$$

的一个算法.

解决这一问题的一种算法是:

S1 $S\leftarrow1$;

S2 $I\leftarrow3$;

S3 $S\leftarrow S\times I$;

S4 $I\leftarrow I+2$;

S5 如果 $I\leqslant99$,那么转 S3;

S6 输出 S.

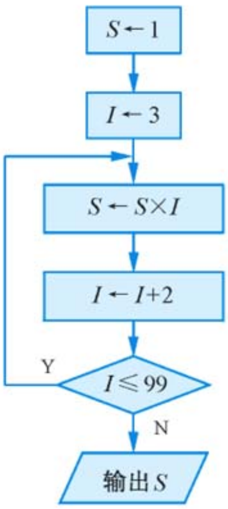


图 5-3-5

画出相应的流程图,如图 5-3-5 所示. 从流程图可以看出这是一个循环结构,我们可以运用循环语句(cycle statement)来实现上述过程.

当循环的次数已经确定,可用"For"语句表示. "For"语句的一般形式为:

```
For I from"初值" to "终值" step "步长" … End for
```

上述问题用循环语句表示为:

```
S ← 1
For I from 3 to 99 step 2
  S ← S × I
End for
Print S
```

上面"For"和"End for"之间缩进的步骤称为循环体. 如果省略"step 2",那么重复循环时,I 的值每次增加 1.

如果我们将上面的问题改为

$$1\times3\times5\times7\times\cdots\times\_\_\_\_\_>10\,000,$$

那么,如何寻找满足条件的最小整数呢?

算法步骤如下:

S1 $S\leftarrow1$;

S2 $I\leftarrow3$;

S3 如果 $S\leqslant10\,000$,那么 $S\leftarrow S\times I$,$I\leftarrow I+2$,重复 S3;

S4 输出 I.

当循环次数不能确定时，可用“While”语句来实现循环.“While”语句的一般形式为

```
While A
…
End while
```

其中 A 表示判断执行循环的条件.

上述问题用“While”语句可描述如下：

```
S ← 1
I ← 3
While S ≤ 10 000
    S ← S × I
    I ← I + 2
End while
Print I
```

先检验条件“$S\leqslant 10\,000$”是否成立. 如果“$S\leqslant 10\,000$”为真，则重复“$S\leftarrow S\times I$, $I\leftarrow I+2$”，只有当条件“$S\leqslant 10\,000$”为假时，才结束重复.

上面“While”和“End while”之间缩进的步骤称为循环体.

“While”语句的特点是“前测试”，即先判断，后执行. 若初始条件不成立，则一次也不执行循环体中的内容. 任何一种需要重复处理的问题都可以用这种前测试循环来实现.

例 4 抛掷一枚硬币时，既可能出现正面，也可能出现反面，预先作出确定的判断是不可能的，但是假如硬币质量均匀，那么当抛掷次数很多时，出现正面的频率应接近于 50%. 试设计一个循环语句模拟抛掷硬币的过程，并计算抛掷中出现正面的频率.

分析 抛掷硬币的过程实际上是一个不断重复地做同一件事情的过程，利用循环语句，我们容易在计算机上模拟这一过程.

在程序语言中，有一个随机函数“Rnd”，它能产生 0 与 1 之间的随机数. 这样，我们可用大于 0.5 的随机数表示出现正面，不大于 0.5 的随机数表示出现反面.

解 本题算法的伪代码如下：

```
s ← 0
Read n
For i from 1 to n
  If Rnd > 0.5 Then s ← s + 1
Next i
Print 出现正面的频率为 s/n
```

统计正面出现的次数.

EXCEL

VBA(Visual Basic for Application)是 Excel 自带的一种程序设计语言,它具有一般程序设计语言所具有的功能,可由手工写入或宏记录器两种方式生成.

使用 VBA 宏记录器无需亲自写 VBA 的代码,在计算机内会自动生成 VBA 的代码.你只要打开宏记录器,做一次你所需要的操作,例如画一个经常要用的表格,其每一步操作它都会用代码记录下来,操作完成后,保存为一个叫宏的文件.下次再做同样的事,你只要执行该文件,就可自动画出已设计好的表格.当然,如果没有相关记录,就要靠人工编写 VBA 程序语言来弥补.

如图 5-3-6 所示,在 Excel 工作表中,选择“工具/宏/Visual Basic 编辑器”.在 VB 编辑器窗口中选择“工具/宏”,在弹出的对话框中,在“宏名称”栏内输入宏的名称,如“抛掷硬币”,点击“创建”,出现宏主体语句 Sub 和 End Sub,输入你的程序后按 F5 即可运行,如不满意,可随时修改.

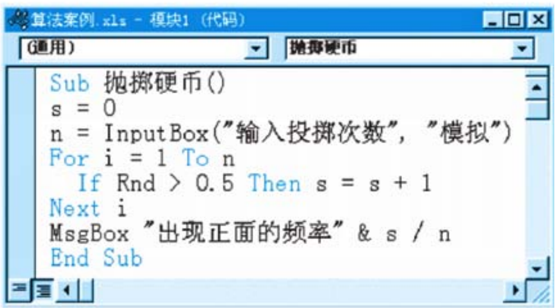


图 5-3-6

当抛掷次数为 10 000 时,可得出现正面的频率为 0.494 4(你的模拟结果可能和这里不同).

练 习

1. 在第 5.2 节开头,我们曾研究过问题 $1+2+3+\cdots+\_\_\_\_ > 2\,004$,试用“While”语句描述这一问题的算法过程.

2. 2000 年我国人口数约为 13 亿.如果每年的人口自然增长率为 15‰,那么多少年后我国人口将达到或超过 15 亿?

这个问题可通过循环方式计算完成,即每一次在原有的基础上增加 15‰,直到达到或超过 15 亿,再记下循环次数.试用循环语句表示这一过程.

3. 根据第 11 页练习 2 的流程图,用条件语句和循环语句描述该算法过程.

4. 1, 1, 2, 3, 5, 8, 13, …这一列数的规律是:第 1、第 2 个数是 1,从第 3 个数起,该数是其前面 2 个数之和.试用循环语句描述计算这列数中前 20 个数之和的算法.

习题 5.2

感受·理解

1. 下列算法中，最后输出的 a, b, c 各是多少？

$a \leftarrow 3$
$b \leftarrow -5$
$c \leftarrow 6$
$a \leftarrow b$
$b \leftarrow c$
Print a, b, c

（第 1 题）

2. 用算法语句给出用公式法求方程 $x^2-3x-4=0$ 的两个根的算法.

3. 输入 3 个数 a, b, c，如果这 3 个数能作为一个三角形的三边长，则输出 $\frac{1}{2}(a+b+c)$，否则提示重新输入. 试用算法语句表示上述过程.

4. 某班有 50 名学生，现将某科的成绩分为 3 个等级：不低于 80 分为 A，低于 60 分为 C，其余为 B. 试用条件语句表示输出每个学生相应的成绩等级的算法.

5. 写出求所有立方小于 1 000 的正整数的算法，并画出流程图，写出伪代码.

6. 设计一个计算 $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{100}$ 的算法，并画出流程图，写出伪代码.

思考·运用

7. 用循环语句描述求 $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{99}-\frac{1}{100}$ 的算法.

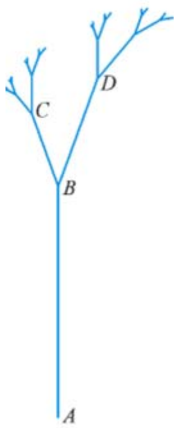
8. 青年歌手大奖赛有 10 名选手参加，并请了 12 名评委. 为了减少极端分数的影响，通常去掉一个最高分和一个最低分后再求平均分. 请用算法语句表示：输入 12 名评委所打的分数 $a_i(i=1, 2, \cdots, 12)$，用函数 $\mathrm{Max}(a_1, a_2, \cdots, a_{12})$ 和 $\mathrm{Min}(a_1, a_2, \cdots, a_{12})$ 分别求出 $a_i(i=1, 2, \cdots, 12)$ 中的最大值和最小值，最后输出该名歌手的成绩.

探究·拓展

9. 通过计算机验证：

(1) 存在自然数 n，使 $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}>10^5$；

(2) 任意给定一个自然数 N，一定存在自然数 n，使 $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}>N$.



（第 10 题）

10. 左边的“树”是按下面的方法作出的：作一条线段 AB，以 B 为一端点分别左拐和右拐 $20°$ 作线段 BC, BD，使 $BC=\frac{1}{3}AB$, $BD=\frac{1}{2}AB$；再分别以 C, D 为端点重复上面的操作，如此连续操作五次. 写出画这棵“树”的流程图，并写出伪代码.（有兴趣的同学，可设计成计算机语言，上机操作）

5.4 算法案例

本节我们选取了若干引人入胜的数学问题，通过对这些数学问题算法的探讨，可以进一步体会算法的思想，提高逻辑思维能力和算法设计水平.

案例 1 设计解决“韩信点兵—孙子问题”的算法.

韩信是秦末汉初的著名军事家.据说有一次汉高祖刘邦在卫士的簇拥下来到练兵场，刘邦问韩信有什么办法，不要逐个报数，就能知道场上士兵的人数.

韩信先令士兵排成 3 列纵队进行操练，结果有 2 人多余；接着他立刻下令将队形改为 5 列纵队，这一改，又多出 3 人；随后他又下令改为 7 列纵队，这一次又剩下 2 人无法成整行.

在场的人都哈哈大笑，以为韩信无法清点出准确人数，不料笑声刚落，韩信便高声报告共有士兵 2 333 人.

众人听了一愣，不知韩信用什么办法这么快就能得出正确结果.

当然，韩信当时是否这样做，已无从查考，但这个故事却说出一个著名的数学问题，即闻名世界的“孙子问题”.

秦九韶（1202～1261），四川安岳人. 1247 年写成名著《数书九章》，其中最重要的成就之一就是“大衍求一术”，即一次同余式组的解法.有兴趣的同学请上网查阅相关介绍.

这种神机妙算，最早出现在我国《算经十书》之一的《孙子算经》中.原文是：“今有物不知其数，三三数之剩二，五五数之剩三，七七数之剩二.问物几何？答曰：二十三.”

自从《孙子算经》中提出这个“物不知数”问题之后，它便引起了人们很大的兴趣.南宋数学家秦九韶对此加以推广，又发现了一种新的算法，叫“大衍求一术”.这种解法后来传入欧洲，欧洲学者发现此解法和高斯的解法本质上是一致的，但比高斯早了 500 余年.

所以，人们将这种问题的通用解法称为“孙子剩余定理”或“中国剩余定理”(Chinese remainder theorem).这个定理在近代数学和电子计算机程序设计中有着广泛的应用.

■ 算法设计思想

“孙子问题”相当于求关于 x，y，z 的不定方程组

$$\begin{cases} m = 3x + 2, \\ m = 5y + 3, \\ m = 7z + 2 \end{cases}$$

的正整数解.

设所求的数为 m，根据题意 m 应同时满足下列三个条件：

> 这里 $\text{int}(x)$ 表示不超过 x 的最大整数，$\text{mod}(a, b)$ 表示 a 除以 b 所得的余数，称 b 为模.

(1) m 被 3 除后余 2，即

$$m-\text{int}(m/3)\times 3=2 \text{ 或 } \text{mod}(m, 3)=2;$$

(2) m 被 5 除后余 3，即

$$m-\text{int}(m/5)\times 5=3 \text{ 或 } \text{mod}(m, 5)=3;$$

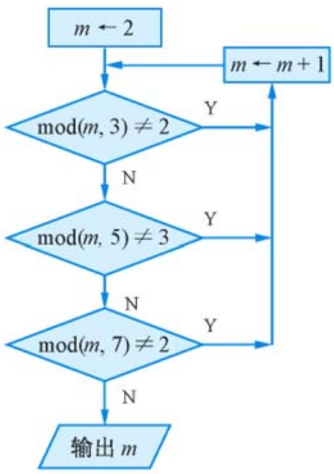
(3) m 被 7 除后余 2，即

$$m-\text{int}(m/7)\times 7=2 \text{ 或 } \text{mod}(m, 7)=2.$$

首先，让 $m=2$ 开始检验条件，若三个条件中有任何一个不满足，则 m 递增 1，一直到 m 同时满足三个条件为止.

■ 流程图与伪代码

> 若 $\text{mod}(m, 3)\neq 2$ 为假，执行下一语句，否则 m 递增 1. 右侧虚线框内的“10，20，…，90”表示相应语句的标号.




```
10  m ← 2
20  If mod(m, 3) ≠ 2 then 70
30  If mod(m, 5) ≠ 3 then 70
40  If mod(m, 7) ≠ 2 then 70
50  Print m
60  Goto 90
70  m ← m+1
80  Goto 20
90  End if
```

图 5-4-1

EXCEL

如图 5-4-2，在 VB 编辑器窗口中创建名称为“孙子问题”的宏，输入如下程序后按 F5 运行，可得不定方程的一个解为 23.

> 框中“<>”表示不等于.

图 5-4-2

案例 2 写出求两个正整数 a，b $(a>b)$ 的最大公约数的一个算法.

公元前 3 世纪，欧几里得在《原本》第七篇中介绍了求两个正整数 a，b $(a>b)$ 的最大公约数的方法，求出一列数：

$$a,\ b,\ r_1,\ r_2,\ \cdots,\ r_{n-1},\ r_n,\ 0.$$

这列数从第三项开始，每项都是前两项相除所得的余数，余数为 0 的前一项 r_n 即是 a 和 b 的最大公约数(greatest common factor). 这种方法称为“欧几里得辗转相除法”(Euclid algorithm).

例如，求 $a=204$ 与 $b=85$ 的最大公约数的步骤为：

$204\div 85$，余数 r_1 为 34，所以

$$204=85\times 2+34;$$

$85\div 34$，余数 r_2 为 17，所以

$$85=34\times 2+17;$$

$34\div 17$，余数为 0，所以

$$34=17\times 2.$$

因此，204 与 85 的最大公约数是 $r_2=17$.

欧几里得(Euclid，约公元前 330～公元前 275)，古希腊数学家，以其所著的《原本》闻名于世.

■ 算法设计思想

欧几里得辗转相除法找出 a，b 的最大公约数的步骤是：计算出 $a\div b$ 的余数 r，若 $r=0$，则 b 为 a，b 的最大公约数；若 $r\neq 0$，则把前面的除数 b 作为新的被除数，把余数 r 作为新的除数，继续运算，直到余数为 0，此时的除数即为正整数 a，b 的最大公约数.

求 a，b 最大公约数的算法为：

S1 输入两个正整数 a，b $(a>b)$；

S2 $r\leftarrow a\div b$ 的余数；

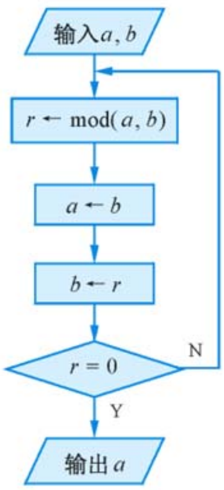
S3 $a\leftarrow b$，$b\leftarrow r$；

S4 若 $r=0$，则输出最大公约数 a；若 $r\neq 0$，则转 S_2.

■ 流程图与伪代码

利用函数 mod(a，b)，得到 a 除以 b 的余数.

当余数 r 为 0 时，输出两个自然数的最大公约数；当余数不为 0 时，重复执行.




```
10  Read a, b
20  r ← mod(a, b)
30  a ← b
40  b ← r
50  If r ≠ 0, then 80
60  Print a
70  Goto 90
80  Goto 20
90  End if
```

图 5-4-3

EXCEL

创建名称为“最大公约数”的宏，输入程序后按 F5 运行，先后输入两个自然数 204 和 85，可得这两个数的最大公约数为 17（如图 5－4－4）.

图 5－4－4

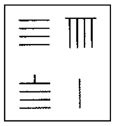
阅　读

辗转相除与更相减损

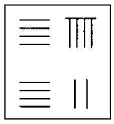
我国的《九章算术》卷一方田章中的“更相减损术”，与辗转相除法非常相似，书中说道：“约分术曰：可半者半之，不可半者付置分母分子之数，以少减多，更相减损，求其等也. 以等数约之.”意思是：如果分母、分子都是偶数（可半者），那么先除以 2；如果不全是偶数，便将分子分母互减，以少减多，直到得出最大公约数为止；用最大公约数去约分子分母，便可使分数最简.

例如，求 49 与 91 的最大公约数，用更相减损是

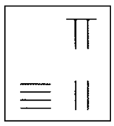
第一步　将 49，91 上下对齐；

49
91


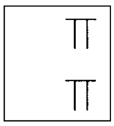
第二步　从 91 中减去 49 余 42；

49
42


第三步　从 49 中减去 42 余 7；

7
42


第四步　从 42 中五次减去 7 出现上下两个相等的数 7.

7
7


陕西旬阳出土的唐代象牙算筹

于是，这个 7 就是“等数”，也就是最大公约数. 由于我国古代用的是筹算，数目都是用算筹排成的，相减时只需取走或重新摆一下算筹即可. 因此上述运算过程虽然是四步，摆出了四个筹式，但实际操作时只需在一个筹式上连续进行就可以了.

可以看出，更相减损正是“欧几里得辗转相除法”.

案例 3 写出用区间二分法求方程 $x^3-x-1=0$ 在区间$[1, 1.5]$内的一个近似解(误差不超过 0.001)的一个算法.

■ 算法设计思想

如图 5-4-5,如果估计出方程 $f(x)=0$ 在某区间$[a, b]$内有一个根 x^*,就能用二分搜索求得符合误差限制 c 的近似解.

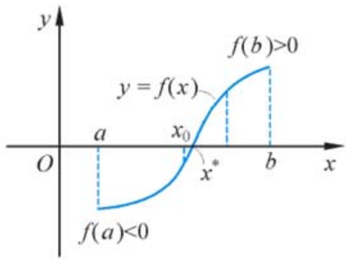


图 5-4-5

它的解法步骤可表示为:

S1 取$[a, b]$的中点 $x_0=\frac{1}{2}(a+b)$,将区间一分为二;

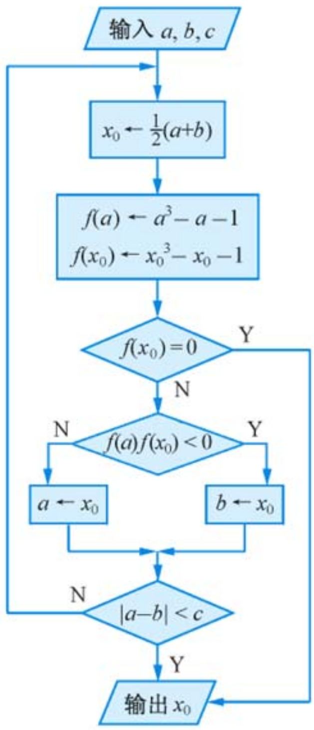
S2 若 $f(x_0)=0$,则 x_0 就是方程的根;否则判别根 x^* 在 x_0 的左侧还是右侧:

若 $f(a)\cdot f(x_0)>0$,则 $x^*\in(x_0, b)$,以 x_0 代替 a;

若 $f(a)\cdot f(x_0)<0$,则 $x^*\in(a, x_0)$,以 x_0 代替 b;

S3 若 $|a-b|<c$,计算终止,此时 $x^*\approx x_0$,否则转 S1.

■ 流程图与伪代码




```
10   Read a, b, c
20   x0 ← (a + b)/2
30   f(a) ← a^3 − a − 1
40   f(x0) ← x0^3 − x0 − 1
50   If f(x0) = 0 then Goto 120
60   If f(a)f(x0) < 0 then
70        b ← x0
80   Else
90        a ← x0
100 End if
110 If | a − b | ≥ c then Goto 20
120 Print x0
```

图 5-4-6

EXCEL

创建名称为“二分法”的宏，输入程序后按 F5 运行，依次输入区间左、右端点值 1，1.5，再输入误差限制 0.001，得方程的近似解为 1.325 195 312 5，如图 5－4－7 所示.

函数 Val( ) 的作用是将字符串转换为数值，可提高运算的正确性和精度.

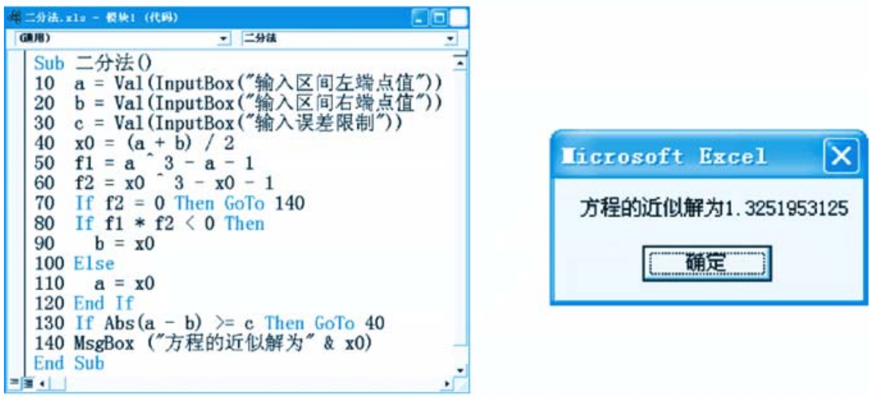


图 5－4－7

练　习

1. 下面一段伪代码的目的是(　　).

```
10  Read x, y
20  m ← x
30  n ← y
40  If m/n=int(m/n) Then Goto 90
50  c ← m−int(m/n)×n
60  m ← n
70  n ← c
80  Goto 40
90  Print n
```

(第 1 题)

A. 求 x，y 的最小公倍数　　　B. 求 x，y 的最大公约数

C. 求 x 被 y 整除的商　　　D. 求 y 除以 x 的余数

2. 在直角坐标系中作出函数 $y=2^x$ 和 $y=x-4$ 的图象，根据图象判断方程 $2^x=x-4$ 的解的范围，再用二分法求这个方程的近似解(误差不超过 0.001)，并写出这个算法的伪代码，画出流程图.(有兴趣的学生可以上机操作)

习题 5.3

感受・理解

1. 一种放射性物质不断变化为其他物质，每经过一年剩留的质量约为原来的 84%. 那么，约经过多少年，剩留的质量是原来的一半？试写出运用二分法计算这个近似解的伪代码.
2. 设计计算两个正整数 a，b 的最小公倍数的算法.

思考・运用

3. 我国古代劳动人民对不定方程的研究作出过重要贡献，其中《张丘建算经》中的“百鸡问题”就是一个很有影响的不定方程问题：今有鸡翁一值钱五，鸡母一值钱三，鸡雏三值钱一. 凡百钱买百只，问鸡翁、母、雏各几何？

   其意思是：一只公鸡的价格是 5 钱，一只母鸡的价格是 3 钱，三只小鸡的价格是 1 钱. 想用 100 钱买 100 只鸡，问公鸡、母鸡、小鸡各可买几只. 设 x，y，z 分别代表公鸡、母鸡、小鸡的只数，我们可以大致确定 x，y，z 的取值范围：若 100 钱全买公鸡，则最多可买 20 只，即 x 的取值范围是 0～20；若 100 钱全买母鸡，则最多可买 33 只，即 y 的取值范围是 0～33；当 x，y 在各自的范围内确定后，则小鸡的只数 $z=100-x-y$ 也就确定了.

   根据上述算法思想，画出求解的流程图，并写出相应的伪代码.
4. 设计求解不定方程 $x_1+2x_2+3x_3+\cdots+11x_{11}+12x_{12}=0$ $(x_1, x_2, \cdots, x_{12}\in\{-1, 1\})$ 的一个算法.（提示：可用循环语句或条件语句）

阅　读

莱布尼茨(G. W. Leibniz，1646～1716)，德国自然科学家、哲学家、数学家.

二进制数・计算机

二进制记数法的思想源远流长，我国古代很早就有研究，在《易经》上就讲到两仪，即一黑一白阴阳互补的两条鱼. 以后，在两仪之上形成了八卦.《易经》中关于两仪及演变的叙述可以看成是二进制应用的萌芽.

德国数学家莱布尼茨 1679 年撰写的《二进制算术》，使他成为二进位数制的发明人. 二进制在现代被应用于计算机设计，但莱布尼茨本人并没有将它用到自己的计算机上. 莱布尼茨后来发现他的二进制可以给中国古老的六十四卦易图一个很好的数学解释，他是通过他的朋友、法国传教士白晋(F. J. Bouvet)得到六十四卦易图的. 莱布尼茨高兴地说：“可以让我加入中国籍了吧！”

莱布尼茨于 1661 年进入莱比锡大学学习，除了学习法律以外，还刻苦研究哲学和数学，他与牛顿几乎同时创立了微积分，在帕斯卡(Blaise Pascal，1623～1662)加法机(加减法)的基础上，他还研制成功能够进行加、减、乘、除和开平方等运算的机械齿轮计算机，并于 1673 年在英国伦敦皇家学会上作了表演.

莱布尼茨发明的计算机

1946 年，世界上第一台电子计算机 ENIAC(埃尼阿克)诞生，是科学技术发展史上一座新的里程碑. 但是它还不完善. 在计算机之父冯・诺伊曼的积极参与和研究之后，很快提出了改进意见. 其中主要的两条对后来计算机科学技术的发展产生了深远影响：第一，用二进制替代原来的十进制(decimal system)，这样大大减少了元器件数量，提高了运算速度；第二，存储程序，就是把程序像数据一样放在计算机内部的存储器中，这也就是后人所说的冯・诺伊曼计算机体系结构. 此后，电子计算机在短短五十多年的时间里得到了飞速发展，成为信息时代的骄子.

冯・诺伊曼(John von Neumann，1903～1957)，美籍匈牙利数学家，对计算机科学、计算机技术和数值分析等作出了开拓性的工作.

计算机为什么要采用二进制呢?

第一，因为二进制只有 0 和 1 两个数字，要得到表示两种不同稳定状态的电子器件很容易，而且制造简单，可靠性高. 例如，电位的高与低，电容的充电与放电，晶体管的导通与截止等.

第二，在各种记数法中，二进制算法逻辑简单，有布尔逻辑代数作理论依据，简单的运算规则使得机器内部的操作也变得简单.

二进制加法法则只有 4 条：

$$0+0=0,\ 0+1=1,$$
$$1+0=1,\ 1+1=10,$$

而十进制加法法则从 $0+0=0$ 到 $9+9=18$，需要 100 条法则.

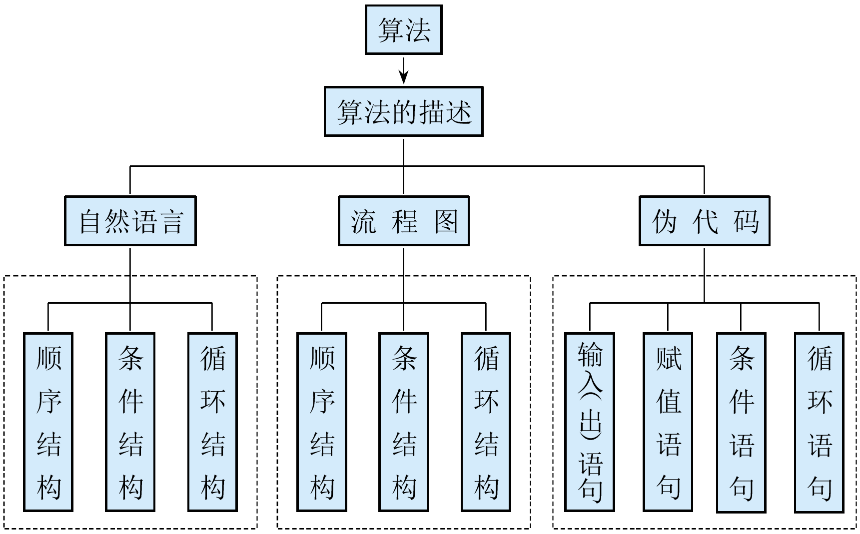
二进制的乘法法则也很简单：

$$0\times0=0,\ 0\times1=0,$$
$$1\times0=0,\ 1\times1=1,$$

而十进制的乘法法则要由一张“九九表”来规定，比较复杂.

本章回顾

本章通过实例介绍了算法的含义，重点研究了在解决问题的过程中如何设计算法，如何根据算法画出流程图，并在此基础上逐步学会运用基本算法语句来表示算法过程.



算法的基本思想就是探求解决问题的一般性方法，并将解决问题的步骤用具体化、程序化的语言加以表述.

描述算法的方式经历了从简单的自然语言向高级的计算机程序语言的发展过程. 自然语言通俗易懂，直接明了；流程图直观形象，能体现算法过程的结构特征；伪代码将流程图的各结构用接近计算机程序语言的算法语句进行表述，为编制计算机程序提供了便利.

复习题

感受·理解

1. 已知⊙O，写出求作⊙O的圆心的一个算法.

2. 已知 $a, b \in \mathbf{N}^*$，且 $a+b=10$，设计一个算法，求出使 ab 取最大值时的 a，b 的值.

3. 已知函数 $f(x)=\dfrac{1}{x}$，设计算法，分别对 $x=1.1, 1.01, 1.001, \cdots, 1.000\,01$ 时计算 $A=\dfrac{f(x)-f(1)}{x-1}$ 的值.

4. 先用不同的算法计算 $\dfrac{1}{1\times 2}+\dfrac{1}{2\times 3}+\dfrac{1}{3\times 4}+\cdots+\dfrac{1}{99\times 100}$，再比较其优劣.

5. 已知△ABC中，$AB=c$，$BC=a$，$\angle B=\alpha$，试写出作△ABC的一个算法.

6. 用条件语句表示：输入 x 的值，通过 $y=\begin{cases}-2x-4, & x\in(-\infty, -2], \\ \sqrt{x+2}, & x\in(-2, 2), \\ 2^{x-1}, & x\in[2, +\infty)\end{cases}$ 计算 y 的值.

7. 根据如图所示的伪代码，可知输出的结果 S 为(　　).

A. 17　　B. 19　　C. 21　　D. 23

```
10  I ← 1
20  While I < 8
30     I ← I + 2
40     S ← 2I + 3
50  End while
60  Print S
```

（第 7 题）

8. 写出求 $a_1, a_2, \cdots, a_{100}$ 中最小数的一个算法.

思考·运用

9. 判断某年份是否为闰年，要看此年份数能否被 4 整除. 如果不能被 4 整除，则是平年，2 月是 28 天；若能被 4 整除，但不能被 100 整除，则该年为闰年，2 月是 29 天；若能被 4 整除，又能被 100 整除，还要看能否被 400 整除，若能则为闰年，否则也为平年. 画出上述算法的流程图，并写出伪代码.

10. 函数 $y=x^2$ 与 $y=2^x$ 有三个交点 (x_1, y_1)，$(2, 4)$，$(4, 16)$，其中 $-1<x_1<0$. 试用二分法求出 x_1 的近似值（误差不超过 0.01）.

11. 要判断一个数 x 是否为质数，我们可以把它分别除以从 2 到 $x-1$ 的每一个整数，如果都除不尽，则 x 为质数. 要判断 a 是否能被 b 整除，只要看 $\dfrac{a}{b}$ 是否等于 $\mathrm{int}\left(\dfrac{a}{b}\right)$，若相等则能整除.

下面是寻找 3～100 之内质数的一个算法的伪代码：

```
10  For x from 3 to 100
20     For i from 2 to x-1
30        If int(x/i) = x/i then Goto 10
40     End for
50     Print x
60  End for
```

（第 11 题）

实际上，上述算法的运算次数较多，可以加以改进. 首先，偶数(2 除外)不可能是质数，因此第 1 行的步长可改为 2. 其次，第 2 行中的 $x-1$ 可改为 $\sqrt{x}$(为什么?).

写出改进后的伪代码，你有寻找质数更好的方法吗?

12. 收集身边的实际问题，设计相应的算法，并与同学交流.

探究·拓展

13. 满足方程 $a^2+b^2=c^2$ 的一组正整数称为勾股数或商高数. 设计计算某一范围内的勾股数的算法(给出算法步骤、画出流程图).

只有将数学应用于社会科学的研究之后，才能使得文明社会的发展成为可控制的现实.

——怀　特

灯泡厂要了解生产的灯泡的使用寿命，需要将所有灯泡逐一测试吗？

保险公司为对人寿保险制定适当的赔偿标准，需要了解人口的平均寿命，怎样获得相关数据？

国际奥委会 2003 年 6 月 29 日决定，2008 年北京奥运会的举办日期将比原定日期推迟两周，改在 8 月 8 日至 8 月 24 日举行. 原因是 7 月末 8 月初北京地区的气温高于 8 月中下旬. 这一结论是如何得到的呢？

……

现实生活中存在大量类似的问题，如国家经济运行状况、人口普查、市场调查及天气预报等.

- 如何科学、合理地收集数据？
- 怎样分析和研究数据，对一般情况作出估计？

6.1 抽样方法

我们已经学习过一些统计知识，了解了统计的基本思想方法是用样本估计总体，即当总体容量很大或检测过程具有一定的破坏性时，不直接去研究总体，而是通过从总体中抽取一个样本，根据样本的情况去估计总体的相应情况. 如气象工作者对过去北京若干年7月下旬到8月下旬的日最高气温进行抽样研究，从而得到对北京一般年份7月25日到8月10日与8月8日到8月24日两个时段的高温分布状况的估计，作出合理的决策.

● 如何科学地进行抽样呢？

6.1.1 简单随机抽样

1. 抽签法

● 为了了解高一(1)班50名学生的视力状况，从中抽取10名学生进行检查. 如何抽取呢？

通常使用抽签法，方法是：将50名学生从1到50进行编号，再制作1到50的50个号签，把50个号签集中在一起并充分搅匀，最后随机地从中抽10个号签. 对编号与抽中的号签的号码相一致的学生进行视力检查.

一般地，用抽签法从个体个数为 N 的总体中抽取一个容量为 k 的样本的步骤为：

(1) 将总体中的所有个体编号(号码可以从1到 N)；

(2) 将1到 N 这 N 个号码写在形状、大小相同的号签上(号签可以用小球、卡片、纸条等制作)；

(3) 将号签放在同一箱中，并搅拌均匀；

(4) 从箱中每次抽出1个号签，并记录其编号，连续抽取 k 次；

(5) 从总体中将与抽到的签的编号相一致的个体取出.

在学习了求概率的方法后，可以说明用抽签法能使每个个体被抽中的概率相等.

这样就得到一个容量为 k 的样本. 对个体编号时，也可以利用已有的编号，如从全班学生中抽取样本时，利用学生的学号作为编号；对某场电影的观众进行抽样调查时，利用观众的座位号作为编号等.

抽签法简单易行，适用于总体中个体数不多的情形.

2. 随机数表法

用抽签法抽取样本时，编号的过程有时可以省略(如用已有的编号)，但制签的过程就难以省去了，而且制签也比较麻烦. 如何简化制签的过程呢？

一个有效的办法是制作一个表，其中的每个数都是用随机方法产生的，这样的表称为随机数表(random number table). 于是，我们只要按一定的规则到随机数表中选取号码就可以了. 这种抽样方法叫做随机数表法.

链　接

随机数表的制作

随机数表是人们根据需要编制出来的，由 0，1，2，3，4，5，6，7，8，9 十个数字组成，表中每一个数字都是用随机方法产生的(称为“随机数”). 随机数的产生方法主要有抽签法、抛掷骰子法和计算机生成法.

(1) 抽签法：用 0，1，2，3，4，5，6，7，8，9 十个数字做十个签，放入一个箱中并搅拌均匀，再从箱中每次抽取一个签并记下签的数码，再放回箱中，如此重复进行下去即可得到一个随机数表.

同桌的两位同学相互协作，编制一张随机数表.

如果需要两位数表，则将所得的各个数码按顺序两两连在一起. 若要三位数表，就三三连在一起，如 012，321，249，460，634，105，….

(2) 抛掷骰子法：如图 6－1－1，在一个正 20 面体的各面写上 0～9 这十个数字(相对的两个面上的数字相同)，这样就得到一个产生 0～9 的随机数的骰子. 不断抛掷这个骰子，并逐一记下朝上一面(与地面或桌面平行)的数字，就能按顺序排成一个随机数表.

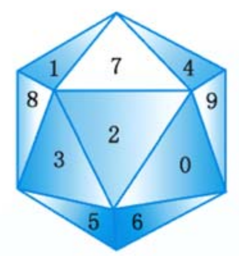


图 6－1－1

(3) 计算机生成法：利用随机函数或随机数发生器让计算机自动生成随机数表.

EXCEL & CALCULATOR

用 Excel 可以方便地产生所需的随机数，下面介绍两种常用的方法.

(1) 利用随机函数生成随机数. 在单元格 C1 内输入随机函数“＝RAND()”，就能得到一个 0 与 1 之间的随机数，拖曳 C1 的填充柄，便可产生不同的随机数(图 6－1－2 中第 C 列). Excel 有 65 536 行，利用 Excel 的自动填充功能，能快速生成大量的随机数.

如果要生成从 0 到 99 的随机数，且随机数为整数，可在单元格内输入“＝INT(100 * RAND())”(图 6－1－2 中第 D 列).

(2) 利用随机数发生器. 依次选择“工具/数据分析/随机数发生

器”后，弹出如图 6－1－3 所示的对话框. 它能根据你的要求产生所需的随机数. 图 6－1－3 中的设置，表示在第 A 列生成 20 个 0 与 1 之间的随机数.

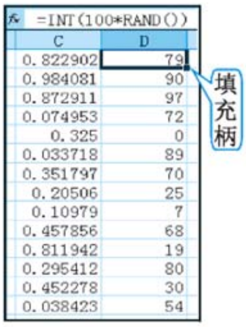


图 6-1-2

图 6-1-3

利用科学型计算器可以产生随机数.如按 SHIFT Ran# = 键可产生 0 与 1 之间的一个随机数,之后每按一次 = 键,计算器就产生一个随机数.

如果要生成某个范围内的随机整数(如从 0 到 99),可按下述方法操作:

(1) 按 MODE MODE MODE 1 0,使计算器以整数的形式显示;

(2) 按 100 SHIFT Ran# =,可产生 0 与 99 之间的一个随机数,且随机数为整数.

之后每按一次 =,计算器就产生一个 0 与 99 之间的随机数.

下面我们用随机数表法求解本节开头的问题.

(1) 对 50 个同学进行编号,编号分别为 01, 02, 03, …, 50;

随机数表见附录 1.

(2) 在随机数表中随机地确定一个数作为开始,如第 8 行第 29 列的数 7 开始.为便于说明,我们将附表中的第 6 行至第 10 行摘录如下:

			第 29 列		
	16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
	84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76
第 8 行	63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79
	33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54
	57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28

(3) 从数 7 开始向右读下去,每次读两位,凡不在 01 到 50 中的数跳过去不读,遇到已经读过的数也跳过去,便可依次得到

12, 07, 44, 39, 38, 33, 21, 34, 29, 42

这 10 个号码,就是所要抽取的 10 个样本个体的号码.

将总体中的 N 个个体编号时可以从 0 开始,例如当 $N=100$ 时,

编号可以是00，01，02，…，99.这样，总体中的所有个体均可用两位数字号码表示，便于使用随机数表.

例如，可以从抛掷一根大头针落在随机数表上针尖所指的数开始.

当随机地选定开始的数后，读数的方向可以向右，也可以向左、向上、向下等.

由此可见，用随机数表法抽取样本的步骤是：

(1) 对总体中的个体进行编号(每个号码位数一致)；

(2) 在随机数表中任选一个数作为开始；

(3) 从选定的数开始按一定的方向读下去，得到的数码若不在编号中，则跳过；若在编号中，则取出；如果得到的号码前面已经取出，也跳过；如此继续下去，直到取满为止；

(4) 根据选定的号码抽取样本.

从个体数为 N 的总体中不重复地取出 n 个个体($n<N$)，每个个体都有相同的机会被取到. 这样的抽样方法称为简单随机抽样.

抽签法和随机数表法都是简单随机抽样(simple random sampling).

练　习

1. 一个学生在一次知识竞赛中要回答的 8 道题是这样产生的：从 15 道历史题中随机抽出 3 道，从 20 道地理题中随机抽出 3 道，从 12 道生物题中随机抽出 2 道. 试用抽签法确定这个学生所要回答的三门学科的题的序号(历史题编号分别为 1，2，…，15，地理题编号分别为 16，17，…，35，生物题编号分别为 36，37，…，47).

2. 从 100 件电子产品中抽取一个容量为 25 的样本进行检测，试用随机数表法抽取样本.

3. 假设一个总体有 5 个元素，分别记为 a，b，c，d，e，从中采用不重复抽取样本的方法，抽取一个容量为 2 的样本，样本共有多少个？写出全部可能的样本.

4. 某学校高一年级共有 200 名学生，为了了解这些学生的身高状况，从中抽取一个容量为 15 的样本.

6.1.2　系统抽样

● 某校高一年级共有 20 个班，每班有 50 名学生. 为了了解高一学生的视力状况，从这 1 000 人中抽取一个容量为 100 的样本进行检查，应该怎样抽样？

通常先将各班学生平均分成 5 组，再在第一组(1 到 10 号学生)中，用抽签法抽取一个，然后按照“逐次加 10(每组中个体个数)”的规则分别确定学号为 11 到 20、21 到 30、31 到 40、41 到 50 的学生代表.

将总体平均分成几个部分，然后按照预先定出的规则，从每个部分中抽取一个个体，得到所需的样本，这样的抽样方法称为系统抽样(systematic sampling).

系统抽样的步骤为：

(1) 采用随机的方式将总体中的个体编号；

(2) 将整个的编号按一定的间隔(设为 k)分段，当 $\frac{N}{n}$(N 为总体中的个体数，n 为样本容量)是整数时，$k=\frac{N}{n}$；当 $\frac{N}{n}$ 不是整数时，从总体中剔除一些个体，使剩下的总体中个体的个数 N' 能被 n 整除，这时 $k=\frac{N'}{n}$，并将剩下的总体重新编号；

(3) 在第一段中用简单随机抽样确定起始的个体编号 l；

(4) 将编号为 l, $l+k$, $l+2k$, …, $l+(n-1)k$ 的个体抽出.

系统抽样也可称为“等距抽样”.

例 1　某单位在岗职工共 624 人，为了调查工人用于上班途中的时间，决定抽取 10%的工人进行调查. 如何采用系统抽样方法完成这一抽样？

分析　因为 624 的 10%约为 62，624 不能被 62 整除，为了保证“等距”分段，应先剔除 4 人.

解　第一步　将 624 名职工用随机方式进行编号；

第二步　从总体中剔除 4 人(剔除方法可用随机数表法)，将剩下的 620 名职工重新编号(分别为 000, 001, 002, …, 619)，并分成 62 段；

第三步　在第一段 000, 001, 002, …, 009 这十个编号中用简单随机抽样确定起始号码 i_0；

第四步　将编号为

$$i_0,\ i_0+10,\ i_0+20,\ \cdots,\ i_0+610$$

的个体抽出，组成样本.

探　究

请你就中学生普遍关心的某一问题，通过网上调查或对本校学生(有条件的话可扩大调查范围)进行抽样调查，了解学生对此问题的看法，并对两种调查方法所得结果进行比较分析(图 6－1－4).

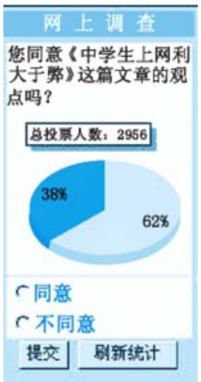


图 6－1－4

练　习

1. 为了了解参加一次知识竞赛的 1 252 名学生的成绩，决定采用系统抽样的方法抽取一个容量为 50 的样本，那么总体中应随机剔除个体的数目是(　　).

   A. 2　　　B. 3　　　C. 4　　　D. 5

2. 要从 1 003 名学生中选取一个容量为 20 的样本，试叙述系统抽样的步骤.
3. 试用系统抽样的方法从你校学生中抽取适当的样本，再对抽出的学生的两臂平展的长度及身高进行测量，分别计算两组数据的平均数.

6.1.3　分层抽样

● 某校高一、高二和高三年级分别有学生 1 000，800 和 700 名，为了了解全校学生的视力情况，欲从中抽取容量为 100 的样本，怎样抽样较为合理？

由于不同年级的学生视力状况有一定的差异，不能在 2 500 名学生中随机抽取 100 名学生，也不宜在三个年级中平均抽取. 为准确反映客观实际，不仅要使每个个体被抽到的概率相等，而且要注意总体中个体的层次性.

一个有效的方法是，使选取的样本中各年级学生所占的比与实际人数占总体人数的比基本相同.

据此，应抽取高一学生 $100\times\frac{1\,000}{2\,500}=40$ 名，高二学生 $100\times\frac{800}{2\,500}=32$ 名，高三学生 $100\times\frac{700}{2\,500}=28$ 名.

一般地，当总体由差异明显的几个部分组成时，为了使样本更客观地反映总体情况，我们常常将总体中的个体按不同的特点分成层次比较分明的几部分，然后按各部分在总体中所占的比实施抽样，这种抽样方法叫**分层抽样**(stratified sampling)，其中所分成的各个部分称为“层”.

分层抽样的步骤是：

若按比例计算所得的个体数不是整数，可作适当的近似处理.

(1) 将总体按一定标准分层；
(2) 计算各层的个体数与总体的个体数的比；
(3) 按各层个体数占总体的个体数的比确定各层应抽取的样本容量；
(4) 在每一层进行抽样(可用简单随机抽样或系统抽样).

例 2　某电视台在因特网上就观众对某一节目的喜爱程度进行调查，参加调查的总人数为 12 000 人，其中持各种态度的人数如表 6-1-1 所示：

表 6-1-1

很喜爱	喜　爱	一　般	不喜爱
2 435	4 567	3 926	1 072

电视台为进一步了解观众的具体想法和意见，打算从中抽取 60 人进行更为详细的调查，应怎样进行抽样？

分析　因为总体中人数较多，所以不宜采用简单随机抽样. 又由于持不同态度的人数差异较大，故也不宜用系统抽样方法，而以分层抽样为妥.

解 可用分层抽样方法，其总体容量为 12 000.

“很喜爱”占 $\frac{2\,435}{12\,000}=\frac{487}{2\,400}$，应取 $60\times\frac{487}{2\,400}\approx 12$ 人；

“喜爱”占 $\frac{4\,567}{12\,000}$，应取 $60\times\frac{4\,567}{12\,000}\approx 23$ 人；

“一般”占 $\frac{3\,926}{12\,000}$，应取 $60\times\frac{3\,926}{12\,000}\approx 20$ 人；

“不喜爱”占 $\frac{1\,072}{12\,000}$，应取 $60\times\frac{1\,072}{12\,000}\approx 5$ 人.

因此，采用分层抽样的方法在“很喜爱”、“喜爱”、“一般”和“不喜爱”的 2 435 人、4 567 人、3 926 人和 1 072 人中分别抽取 12 人、23 人、20 人和5 人.

练 习

1. 在某年有奖明信片销售活动中，规定每 100 万张为一个开奖组，通过随机抽取的方式确定号码的后四位数为 2 709 的为三等奖. 这样确定获奖号码的抽样方法是________.
2. 某公司生产三种型号的轿车，产量分别为 1 200 辆、6 000 辆和 2 000 辆. 为检验该公司的产品质量，现用分层抽样的方法抽取 46 辆进行检验，这三种型号的轿车应分别抽取______、______和______辆.
3. 某所学校有小学部、初中部和高中部，在校小学生、初中生和高中生之比为 5∶2∶3，且已知初中生有 800 人. 现要从这所学校中抽取一个容量为 80 的样本以了解他们对某一问题的看法，应采用什么抽样方法？从小学部、初中部及高中部各抽取多少名？总体上看，平均多少名学生中抽取到一名学生？
4. 将你所在班级的同学按性别分成两个组，分别编号，制成号签，分别放在两个箱子里搅拌均匀，然后按男女生之比各抽出若干个号签，组成两个样本，就他们对某一方面问题的看法进行调查，以比较男女同学对该问题看法上的差异.

以上我们学习了三种抽样方法，这些抽样方法的特点及适用范围可归纳如下：

表 6－1－2

<table>
<tr><th>类别</th><th>特 点</th><th>相互联系</th><th>适用范围</th><th>共同点</th></tr>
<tr><td>简单随机抽样</td><td>从总体中逐个抽取</td><td></td><td>总体中的个体个数较少</td><td rowspan="3">抽样过程中每个个体被抽到的可能性相同</td></tr>
<tr><td>系统抽样</td><td>将总体平均分成几部分，按事先确定的规则分别在各部分中抽取</td><td>在起始部分抽样时，采用简单随机抽样</td><td>总体中的个体个数较多</td></tr>
<tr><td>分层抽样</td><td>将总体分成几层，按各层个体数之比抽取</td><td>各层抽样时采用简单随机抽样或系统抽样</td><td>总体由差异明显的几部分组成</td></tr>
</table>

例3　下列问题中，采用怎样的抽样方法较为合理？

(1) 从10台冰箱中抽取3台进行质量检查；

(2) 某电影院有32排座位，每排有40个座位，座位号为1～40. 有一次报告会坐满了听众，报告会结束以后为听取意见，需留下32名听众进行座谈；

(3) 某学校有160名教职工，其中教师120名，行政人员16名，后勤人员24名. 为了了解教职工对学校在校务公开方面的意见，拟抽取一个容量为20的样本.

分析　(1) 总体容量比较小，用抽签法或随机数表法都很方便.

(2) 总体容量比较大，用抽签法或随机数表法比较麻烦. 由于人员没有明显差异，且刚好32排，每排人数相同，可用系统抽样.

(3) 由于学校各类人员对这一问题的看法可能差异较大，故应采用分层抽样方法.

解　(1) 用抽签法或随机数表法.

(2) 将每排的40个人组成一组，共32组，从第1排至第32排分别为第1～32组，先在第1排用简单随机抽样法抽出一名听众，再将其他各排与此听众座位号相同的听众全部取出.

(3) 总体容量为160，故样本中教师人数应为$20\times\frac{120}{160}=15$名，行政人员人数应为$20\times\frac{16}{160}=2$名，后勤人员人数应为$20\times\frac{24}{160}=3$名.

思　考

某城市两个普通中学分别对自己所在学校12～14岁学生身高进行了抽样统计，发现这两所学校12～14岁学生的平均身高竟相差19 cm，这可能吗？他们在抽样的过程中可能出现了哪些问题？

练　习

1. 某商场想通过检查发票及销售记录的2%来快速估计每月的销售总额，采取如下方法：从某本50张的发票存根中随机抽取一张，如15号，然后按顺序往后将65号、115号、165号、……发票上的销售额组成一个调查样本. 这种抽取样本的方法是(　　).

A. 抽签法　　B. 系统抽样

C. 分层抽样　　D. 随机数表法

2. 某科研机构由行政人员、科技人员和后勤职工三种不同类型的人员组成，现要抽取一个容量为45的样本进行调查. 已知科技人员共有60人，抽入样本的有20人，且行政人员与后勤职工人数之比为2∶3，则此机构的总人数、行政人员、后勤职工人数分别为多少？

3. 一个单位有职工160人，其中业务人员96人，管理人员40人，后勤人员24人. 为了了解职工的某种情况，要从中抽取一个容量为20的样本，按下述三种方法抽取：

① 将160人从1至160编号，用白纸做成有1至160号的签放入箱内拌匀，

然后从中抽取 20 个签，与签号相同的 20 个人被选出；

② 将 160 人从 1 至 160 编号，按编号顺序分成 20 组，每组 8 人，号码分别为 1～8 号、9～16 号、…、153～160 号，先从第 1 组中用抽签法抽出 $k(0<k<9)$ 号，其余组的 $(k+8n)$ 号 $(n=1, 2, \cdots, 19)$ 亦被抽到，如此抽到 20 人；

③ 按 20∶160=1∶8 的比例，从业务人员中抽取 12 人，从管理人员中抽取 5 人，从后勤人员中抽取 3 人，都用随机数表法从各类人员中抽取所需的人数，他们合在一起恰好抽到 20 人.

上述三种抽样方法中，按简单随机抽样法、分层抽样法、系统抽样法的顺序是(   ).

A. ①、②、③　　B. ②、①、③　　C. ①、③、②　　D. ③、①、②

4. 下列抽样中不是系统抽样的是(   ).

A. 从号码为 1～15 的 15 个球中任选 3 个作为样本，先在 1～5 号球中用抽签法抽出 i_0 号，再将号码为 i_0+5，i_0+10 的球也抽出

B. 工厂生产的产品，用传送带将产品送入包装车间的过程中，检验人员从传送带上每隔 5 min 抽一件产品进行检验

C. 搞某市场调查，规定在商店门口随机地询问一个人，直到调查到事先规定的调查人数为止

D. 电影院调查观众的某一指标，邀请每排(每排人数相等)座位号为 14 的观众留下来座谈

习题 6.1

感受·理解

1. 为了了解某市800家企业的管理情况，拟抽取40家企业作为样本.这800家企业中有中外合资企业160家，私营企业320家，国有企业240家，其他性质的企业80家.如何抽取较合理？

2. 用分层抽样的方法从某校学生中抽取一个容量为45的样本，其中高一年级抽20人，高三年级抽10人.已知该校高二年级共有学生300人，求该校学生总数.

3. 某市的4个区共有20 000名学生，且4个区的学生人数之比为3∶2.8∶2.2∶2.现要用分层抽样的方法从所有学生中抽取一个容量为200的样本，那么在这4个区中分别应抽多少名学生？

4. 用适当的方法对你校高一学生的体重和身高进行抽样调查，并将数据收集整理，以备进一步分析.

5. 用系统抽样的方法，对你校高一到高三学生的身高和心率进行抽样调查，并制成表格.

思考·运用

6. 用适当的方法，对你所在学校的学生家长进行抽样调查，将其父亲、母亲的年龄收集整理，并用表格表示出来.

7. 有关方面就男女职工是否同龄退休广泛征求意见，请对你班学生进行抽样调查，了解他们对这一问题的看法.

8. 某校高一年级500名学生中，血型为O型的有200人，A型的有125人，B型的有125人，AB型的有50人.为了研究血型与色弱之间的关系，要从中抽取一个容量为40的样本，应如何抽样？写出AB血型的样本的抽样过程.

9. 举例说明各种抽样方法在实际生活中的应用.

6.2 总体分布的估计

为了了解7月25日至8月24日北京地区的气温分布状况，我们对以往年份这段时间的日最高气温进行抽样，并对得到的数据进行分析. 我们随机抽取近年来北京地区7月25日至8月24日的日最高气温，得到如下样本(单位：℃)：

表6-2-1

7月25日至8月10日	41.9	37.5	35.7	35.4	37.2	38.1	34.7	33.7	33.3
	32.5	34.6	33.0	30.8	31.0	28.6	31.5	28.8	
8月8日至8月24日	28.6	31.5	28.8	33.2	32.5	30.3	30.2	29.8	33.1
	32.8	29.4	25.6	24.7	30.0	30.1	29.5	30.3	

● 怎样通过上表中的数据，分析比较两时间段的高温(≥33℃)状况呢？

6.2.1 频率分布表

上面两样本中的高温天数的频率用下表表示：

表6-2-2

时　间	总天数	高温天数(频数)	频　率
7月25日至8月10日	17	11	0.647
8月8日至8月24日	17	2	0.118

由此表可以发现，近年来，北京地区7月25日至8月10日的高温天气的频率明显高于8月8日至8月24日.

上例说明，当总体很大或不便于获得时，可以用样本的频率分布估计总体的频率分布. 我们把反映总体频率分布的表格称为频率分布表(frequency distribution table).

初中数学已经介绍了历史上所做的抛掷硬币的试验，并获得了正面向上或反面向上的频率分布表. 这类试验只有两种结果，比较简单. 下面研究较为复杂的频率分布表的制作方法.

例 1　从某校高一年级的 1 002 名新生中用系统抽样的方法抽取一个容量为 100 的身高样本，数据如下（单位：cm）. 试作出该样本的频率分布表.

表 6－2－3

168	165	171	167	170	165	170	152	175	174
165	170	168	169	171	166	164	155	164	158
170	155	166	158	155	160	160	164	156	162
160	170	168	164	174	171	165	179	163	172
180	174	173	159	163	172	167	160	164	169
151	168	158	168	176	155	165	165	169	162
177	158	175	165	169	151	163	166	163	167
178	165	158	170	169	159	155	163	153	155
167	163	164	158	168	167	161	162	167	168
161	165	174	156	167	166	162	161	164	166

分析　该组数据中最小值为 151，最大值为 180，它们相差 29，可取区间[150.5，180.5]，并将此区间分成 10 个小区间，每个小区间长度为 3，再统计出每个区间内的频数并计算相应的频率. 我们将整个取值区间的长度称为全距，分成的区间的长度称为组距.

解　(1) 在全部数据中找出最大值 180 和最小值 151，则两者之差为 29，确定全距为 30，决定以组距 3 将区间[150.5，180.5]分成 10 个组；

(2) 从第一组[150.5，153.5)开始，分别统计各组中的频数，再计算各组的频率，并将结果填入下表：

除最后边的区间是闭区间外，其他区间均为左闭右开区间. 称区间的左端点为下组限，右端点为上组限. 此处采用下组限在内而上组限不在内的分组方法，当然也可采用下组限不在内而上组限在内的分组方法.

表 6－2－4

分　组	频数累计	频　数	频　率
[150.5，153.5)	4	4	0.04
[153.5，156.5)	12	8	0.08
[156.5，159.5)	20	8	0.08
[159.5，162.5)	31	11	0.11
[162.5，165.5)	53	22	0.22
[165.5，168.5)	72	19	0.19
[168.5，171.5)	86	14	0.14
[171.5，174.5)	93	7	0.07
[174.5，177.5)	97	4	0.04
[177.5，180.5]	100	3	0.03
合　计		100	1

这张表给出了该身高样本处于各个区间内的人数和频率，由此可估计该校高一学生的身高的分布状况.

一般地，编制频率分布表的步骤如下：

(1) 求全距，决定组数和组距，组距 $=\dfrac{\text{全距}}{\text{组数}}$；

(2) 分组，通常对组内数值所在区间取左闭右开区间，最后一组取闭区间；

(3) 登记频数，计算频率，列出频率分布表.

如果取全距时不利于分组(如不能被组数整除)，可适当增大全距，如在左、右两端各增加适当范围(尽量使两端增加的量相同).

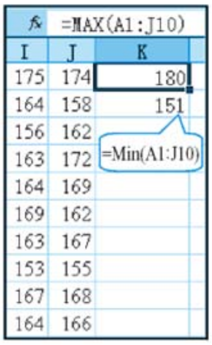
思　考

在编制频率分布表时，分的组数过少或过多各有何利弊？

EXCEL

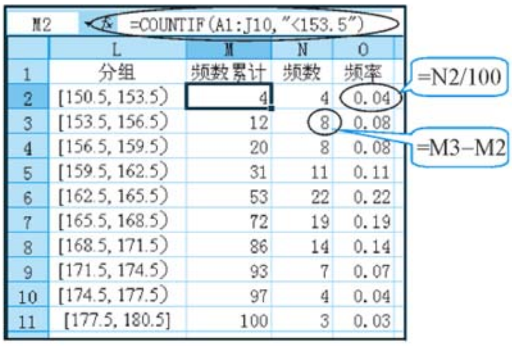
用 Excel 制作频率分布表的操作步骤如下：

(1) 将数据输到 Excel 工作表中(数据区 A1：J10)，在空白单元格插入函数“＝MAX(A1：J10)”和“＝MIN(A1：J10)”可求出数据区域内的最大、最小值分别为 180，151(图 6－2－1)；



I	J	K
175	174	180
164	158	151
156	162	
163	172	
164	169	
169	162	
163	167	
153	155	
167	168	
164	166	

图 6－2－1



	L	M	N	O
1	分组	频数累计	频数	频率
2	[150.5, 153.5)	4	4	0.04
3	[153.5, 156.5)	12	8	0.08
4	[156.5, 159.5)	20	8	0.08
5	[159.5, 162.5)	31	11	0.11
6	[162.5, 165.5)	53	22	0.22
7	[165.5, 168.5)	72	19	0.19
8	[168.5, 171.5)	86	14	0.14
9	[171.5, 174.5)	93	7	0.07
10	[174.5, 177.5)	97	4	0.04
11	[177.5, 180.5]	100	3	0.03

图 6－2－2

(2) 确定全距为 30，组距为 3，将区间[150.5，180.5]分成 10 个组. 用函数 COUNTIF()计算频数累计，如计算身高小于 165.5 的人数，只需输入“＝COUNTIF(A1：J10，"<165.5")”；

COUNTIF()为条件计数函数，前一项为计数对象所在的区域，后一项为计数的条件，结果为符合该条件的数据的个数.

(3) 计算频数和频率. 用该组的累计频数减去上一个组的累计频数即得该组的频数(在单元格 N2 中输入 4，在单元格 N3 内输入“＝M3－M2”，再双击 N3 的填充柄即完成频数的计算). 将频数除以 100 即得频率(单元格 O2 内输入“N2/100”，再双击 N2 的填充柄便完成全部计算)，如图 6－2－2 所示.

探　究

操作 1：

将 1 000 粒黑芝麻与 1 000 粒白芝麻放入一容器中，并搅拌均匀，再用小杯从容器中取出一杯芝麻，计算黑芝麻的频率.

操作 2：

将 1 500 粒黑芝麻与 500 粒白芝麻放入一容器中，并搅拌均匀，再用小杯从容器中取出一杯芝麻，计算黑芝麻的频率.

通过两次试验，你是否有所发现？若有一袋芝麻，由黑、白两种芝麻混合而成，你用什么方法估计其中黑芝麻所占的百分比？

练　习

1. 下面是不同厂家生产的手提式电脑的重量（单位：kg），试列出其频率分布表：

1.9　2.0　2.1　2.4　2.4　2.8　3.2　2.3　1.5　2.6

2.6　1.9　2.4　2.2　1.6　1.7　1.7　1.8　1.8　3.0

2. 一个容量为 20 的数据样本，分组与频数为：[10，20] 2 个、(20，30] 3 个、(30，40] 4 个、(40，50] 5 个、(50，60] 4 个、(60，70] 2 个，则样本数据在区间 $(-\infty, 50]$ 上的可能性为（　　）．

A. 5%　　B. 25%　　C. 50%　　D. 70%

“组中值”为各个组区间的中点的数值．

3. 在一本书中，分组统计 100 个句子中的字数，得出下列结果：字数 1～5 个的 15 句，字数 6～10 个的 27 句，字数 11～15 个的 32 句，字数 16～20 个的 15 句，字数 21～25 个的 8 句，字数 26～30 个的 3 句．请作出字数的频率分布表，并利用组中值对该书中平均每个句子包含的字数作出估计．

自己找一本书，对书中句子所含字数进行抽样估计．

6.2.2　频率分布直方图与折线图

我们还学过一种更为直观地体现数据的分布规律的方法——绘制频数条形图或频率直方图．

例 2　下表是某学校一个星期中收交来的失物件数，请将 5 天中收交来的失物数用条形图表示．

表 6－2－5

星　期	一	二	三	四	五
件　数	6	2	3	5	1
累　计	6	8	11	16	17

解　我们用 Excel 作条形图：

(1) 在 Excel 工作表中输入数据，光标停留在数据区中；

(2) 选择“插入/图表”，在弹出的对话框中点击“柱形图”(bar graph)；

(3) 点击“完成”（如图 6－2－3）．

试作出第 6.2 节开头提出的北京地区 7 月 25 日至 8 月 10 日与 8 月 8 日至 8 月 24 日两时间段的高温天数的频数条形图．

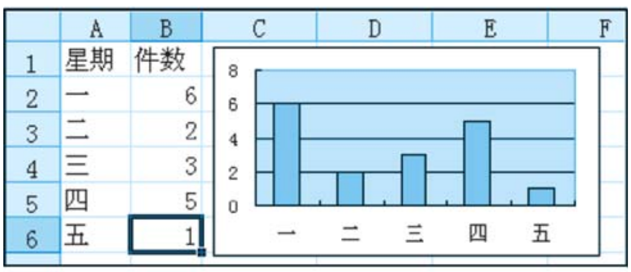


图 6－2－3

我们可以利用直方图反映样本的频率分布规律，这样的直方图称为频率分布直方图(frequency histogram)，简称频率直方图. 下面用例 1 中的数据加以说明.

例 3　作出例 1 中数据的频率分布直方图.

解　(1) 先制作频率分布表，然后作直角坐标系，以横轴表示身高，纵轴表示$\dfrac{频率}{组距}$；

(2) 在横轴上标上 150.5，153.5，156.5，…，180.5 表示的点(为方便起见，起始点 150.5 可适当前移)；

(3) 在上面标出的各点中，分别以连结相邻两点的线段为底作矩形，高等于该组的$\dfrac{频率}{组距}$.

至此，就得到了这组数据的频率分布直方图，如图 6－2－4 所示.

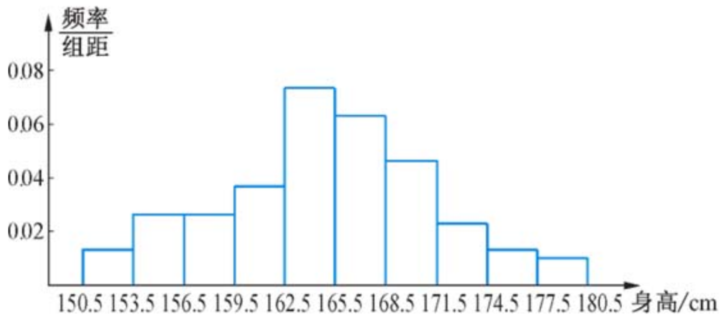


图 6－2－4

一般地，作频率分布直方图的方法为：

把横轴分成若干段，每一线段对应一个组的组距，然后以此线段为底作一矩形，它的高等于该组的$\dfrac{频率}{组距}$，这样得出一系列的矩形，每个矩形的面积恰好是该组上的频率. 这些矩形就构成了频率分布直方图.

频率直方图比频率分布表更直观、形象地反映了样本的分布规律，如在 164 附近达到“峰值”，并具有一定的对称性，这说明这批学生的身高在 164 cm 附近较为集中. 另外还可看出，特别高和特别矮的学生都很少.

如果将频率分布直方图中各相邻的矩形的上底边的中点顺次连结起来，就得到一条折线，我们称这条折线为本组数据的频率折线图(frequency polygon).

例 3 的频率直方图 6－2－4 中，取各相邻矩形的上底边中点并顺次连结，再将矩形的边去除，得频率折线图如图 6－2－5 所示.

取值区间两端点须分别向外延伸半个组距，并取此组距上在 x 轴上的点与折线的首、尾分别相连.

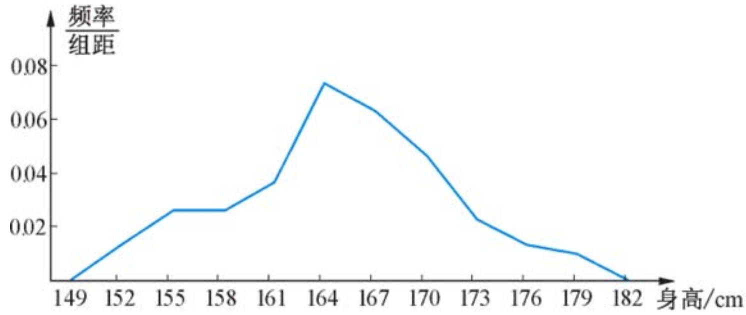


图 6－2－5

思　考

不用频率分布直方图，如何直接作频率折线图？比较几种表示频率分布的方法，各有哪些优点和不足.

频率折线图的优点是它反映了数据的变化趋势. 如果将样本容量取得足够大，分组的组距取得足够小，则这条折线将趋于一条曲线，我们称这一曲线为总体分布的密度曲线.

如例3的频率分布密度曲线可近似地表示为图 6－2－6.

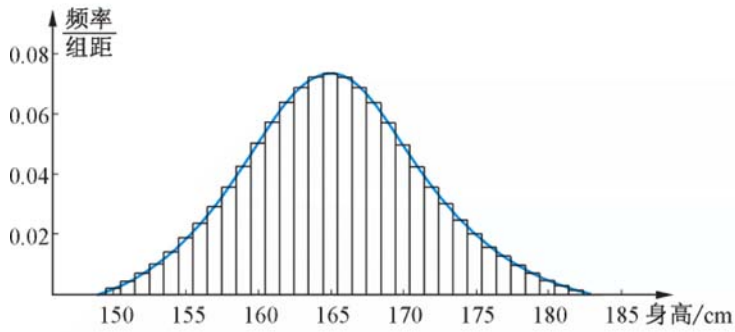


图 6－2－6

例4　为了了解一大片经济林的生长情况，随机测量其中的100株的底部周长，得到如下数据表（长度单位：cm）：

表 6－2－6

135	98	102	110	99	121	110	96	100	103
125	97	117	113	110	92	102	109	104	112
109	124	87	131	97	102	123	104	104	128
105	123	111	103	105	92	114	108	104	102
129	126	97	100	115	111	106	117	104	109
111	89	110	121	80	120	121	104	108	118
129	99	90	99	121	123	107	111	91	100

（续 表）

99	101	116	97	102	108	101	95	107	101
102	108	117	99	118	106	119	97	126	108
123	119	98	121	101	113	102	103	104	108

（1）编制频率分布表；

（2）绘制频率分布直方图；

（3）估计该片经济林中底部周长小于 100 cm 的树木约占多少，周长不小于 120 cm 的树木约占多少.

解 （1）从表中可以看出，这组数据的最大值为 135，最小值为 80，故全距为 55，可将其分为 11 组，组距为 5.

从第 1 组[80，85)开始，将各组的频数、频率和$\frac{频率}{组距}$填入表中.

表 6－2－7

分 组	频 数	频 率	$\frac{频率}{组距}$
[80，85)	1	0.01	0.002
[85，90)	2	0.02	0.004
[90，95)	4	0.04	0.008
[95，100)	14	0.14	0.028
[100，105)	24	0.24	0.048
[105，110)	15	0.15	0.030
[110，115)	12	0.12	0.024
[115，120)	9	0.09	0.018
[120，125)	11	0.11	0.022
[125，130)	6	0.06	0.012
[130，135]	2	0.02	0.004
合 计	100	1	0.2

（2）这组数据的频率直方图如图 6－2－7 所示.

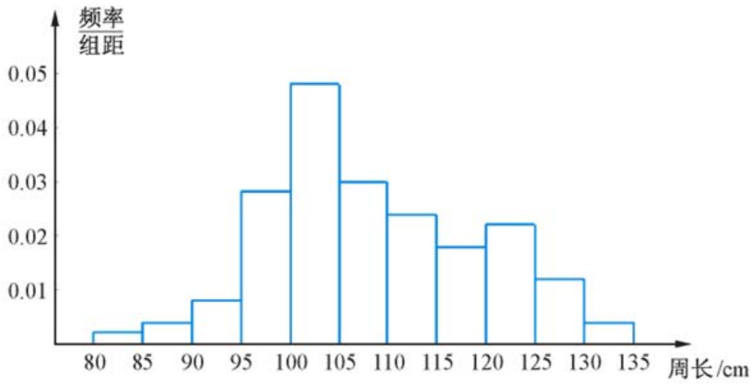


图 6－2－7

(3) 从频率分布表可以看出，该样本中小于100的频率为0.01＋0.02＋0.04＋0.14＝0.21，不小于120的频率为0.11＋0.06＋0.02＝0.19，故可估计该片经济树林中底部周长小于100 cm的树木约占21%，周长不小于120 cm的树木约占19%.

练 习

1. 作出习题6.1第4和第5题中数据的频率分布直方图、频率折线图.

2. 为了了解一批灯泡(共10 000只)的使用寿命，从中抽取了100只进行测试，其使用寿命如下表：

使用寿命/h	500	600	700	800	900	1 000	1 100	1 200	1 300	1 400
只 数	1	4	8	15	20	24	18	7	2	1

(1) 制作频率分布表；
(2) 绘制频率分布直方图；
(3) 根据样本的频率分布，估计使用寿命不低于1 000 h的灯泡约有多少只.

6.2.3 茎叶图

某篮球运动员在某赛季各场比赛的得分情况如下：

12，15，24，25，31，31，36，36，37，39，44，49，50.

● 如何分析该运动员的整体水平及发挥的稳定程度？

初中统计部分曾学过用平均数、众数和中位数反映总体的水平，用方差考察稳定程度.我们还有一种简易的方法，就是将这些数据有条理地列出来，从中观察得分的分布情况.这种方法就是画出该运动员得分的茎叶图(stem and leaf display).

制作茎叶图的方法是：将所有两位数的十位数字作为“茎”，个位数字作为“叶”，茎相同者共用一个茎，茎按从小到大的顺序从上向下列出，共茎的叶一般按从大到小(或从小到大)的顺序同行列出.

上述运动员的得分茎叶图可用图6－2－8来表示.

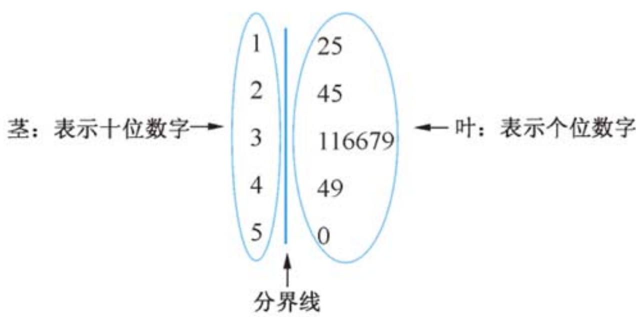


图6－2－8

图 6－2－8 中第一行分界线的左边的“1”表示十位数字，右边的“2”和“5”表示个位数字，这一行说明该运动员的得分为 12 分和 15 分. 同理，第二行说明得分为 24 分和 25 分，第三行说明有两个 31 分，两个 36 分，一个 37 分，一个 39 分，依此类推.

从这张图可以粗略地看出，该运动员平均得分及中位数、众数都在 20 到 40 之间，且分布较对称，集中程度高，说明其发挥比较稳定.

茎叶图既可以分析单组数据，也可以对两组数据进行比较.

例 5 甲、乙两篮球运动员上赛季每场比赛的得分如下，试比较这两位运动员的得分水平.

甲 12，15，24，25，31，31，36，36，37，39，44，49，50.

乙 8，13，14，16，23，26，28，33，38，39，51.

解 画出两人得分的茎叶图，为便于对比分析，可将茎放在中间共用，叶分列左、右两侧(图 6－2－9)：

左侧的叶按从大到小的顺序写，右侧的按从小到大的顺序写. 相同的得分要重复记录，不能遗漏.

甲		乙
	0	8
52	1	346
54	2	368
976611	3	389
94	4	
0	5	1

第二行表示甲得分为 15 分、12 分，乙得分为 13 分、14 分、16 分. 其他各行与此类同.

图 6－2－9

从这个茎叶图可以看出，甲运动员的得分大致对称，平均得分、众数及中位数都是 30 多分. 乙运动员的得分除一个 51 分外，也大致对称，平均得分、众数及中位数都是 20 多分. 因此甲运动员发挥比较稳定，总体得分情况比乙好.

用茎叶图刻画数据有两个优点：一是所有的信息都可以从这个茎叶图中得到，二是茎叶图便于记录和表示. 但茎叶图表示三位数以上的数据时不够方便.

练　习

1. 2002～2003 赛季，一球员在 NBA 某些场次的比赛所得篮板球数分别为 16，6，3，5，12，8，13，6，10，3，19，14，9，7，10，10，9，11，6，11，12，14，8，6，10，5，10，11，13，9，10，10，7，6，11，12，17，4，12，8，10，12，9，15，15，12，13，18，8，16，请制作这些数据的茎叶图.

2. 下面是甲、乙两名运动员某赛季一些场次得分的茎叶图：

甲		乙
	0	8
50	1	247
32	2	199
875421	3	36
944	4	
1	5	2

（第 2 题）

（1）甲、乙两名队员的最高得分各是多少？

（2）哪名运动员的成绩好一些？

习题 6.2

感受・理解

1. 某射手在同一条件下射靶 30 次，其中 6 环或 6 环以下 2 次，7 环 6 次，8 环 7 次，9 环 10 次，10 环 5 次.

（1）列出频率分布表；

（2）根据上述结果，估计射手射中 7～9 环的可能性.

2. 从大量棉花中抽取 50 根棉花纤维，纤维长度（单位：mm）的数据分组及各组的频数为：[22.5，25.5)，3；[25.5，28.5)，8；[28.5，31.5)，9；[31.5，34.5)，11；[34.5，37.5)，10；[37.5，40.5)，5；[40.5，43.5]，4.

（1）列出样本的频率分布表；

（2）画出频率分布直方图；

（3）估计纤维长度小于 36 的百分比.

3. 为了检测某种产品的质量，抽取了一个容量为 100 的样本，数据的分组及频数如下表：

分　组	频　数	频　率
[10.75，10.85)	3	
[10.85，10.95)	9	
[10.95，11.05)	13	
[11.05，11.15)	16	
[11.15，11.25)	26	
[11.25，11.35)	20	
[11.35，11.45)	7	
[11.45，11.55)	4	
[11.55，11.65]	2	
合　计	100	

（1）完成上面的频率分布表；

（2）根据上表画出频率分布直方图；

（3）根据上表和所画的频率分布直方图，估计数据落在[10.95，11.35)范

围内的可能性是百分之几；

(4) 数据小于 11.20 的可能性是百分之几？

可登录 http://www.1088.com.cn/move/003/下载数据.

4. 从规定尺寸为 25.40 mm 的一堆产品中任取 100 件，测得它们的实际尺寸如下(单位：mm)，试列出样本频率分布表，作出频率分布直方图、折线图.

25.39	25.36	25.34	25.42	25.45	25.38	25.39	25.42	25.47	25.35
25.41	25.43	25.44	25.48	25.45	25.43	25.46	25.40	25.51	25.45
25.40	25.39	25.41	25.36	25.38	25.31	25.56	25.43	25.40	25.38
25.37	25.44	25.33	25.46	25.40	25.49	25.34	25.42	25.50	25.37
25.35	25.32	25.45	25.40	25.27	25.43	25.54	25.39	25.45	25.43
25.40	25.43	25.44	25.41	25.53	25.37	25.38	25.24	25.44	25.40
25.36	25.42	25.39	25.46	25.38	25.35	25.31	25.34	25.40	25.36
25.41	25.32	25.38	25.42	25.40	25.33	25.37	25.41	25.49	25.35
25.47	25.34	25.30	25.39	25.36	25.46	25.29	25.40	25.37	25.33
25.40	25.35	25.41	25.37	25.47	25.39	25.42	25.47	25.38	25.39

5. 下面是某市 2003 年 9 月 26 日和 9 月 29 日市区出现堵车的时刻，试列出这两天的堵车时刻的频率分布表和频率分布直方图，并分析该市每天大约在什么时间段是行车高峰期.

9 月 26 日	8:01	8:02	9:30	9:31	9:51	10:24	10:51	
	11:21	15:52	16:30	17:29	17:30	18:04	18:22	
9 月 29 日	8:29	8:32	8:33	9:29	9:58	10:14	10:33	11:43
	14:00	16:08	16:29	16:54	16:55	17:05	18:08	18:09

6. 选一篇英语短文，从它的第一个单词起直到最后一个单词结束，数出各个单词所含字母的个数，并就字母个数列出频率分布表，画出频率分布直方图；也可从电脑的文档里或网上找一篇短文，用计算机查找的方法进行统计.

思考·运用

7. 为了检测某种产品的质量，抽取了一个容量为 40 的样本，检测结果为一等品 8 件，二等品 18 件，三等品 12 件，次品 2 件.

(1) 列出样本的频率分布表；

(2) 画出样本的频率分布直方图；

(3) 根据上述结果，估计这种产品为二等品或三等品的百分率是多少.

8. 在某电脑杂志的一篇文章中，每个句子的字数如下：

10 28 31 17 23 27 18 15 26 24 20 19

36 27 14 25 15 22 11 21 24 27 17 29

在某报纸的一篇文章中，每个句子中所含字的个数如下：

27 39 33 24 28 19 32 41 33 27 35 12

36 41 27 13 22 23 18 46 32 22 18 32

(1) 将这两组数据用茎叶图表示；

(2) 将这两组数据进行比较分析，你能得到什么结论？

9. 上网查阅有关资料，对“福利彩票”的中奖号码进行抽样调查，统计每个数码出现的频率.

阅 读

拉普拉斯(P. S. Laplace, 1749～1827),法国著名的数学家和天文学家,分析概率论的创始人,天体力学的主要奠基人,天体演化学的创立者之一.

男女出生性别比

男女出生的频率都是 0.5 吗? 事实并非如此!

1814 年,法国著名的数学家拉普拉斯在他的著作《概率的哲学讨论》中,记录了一项有趣的统计结果. 他根据伦敦、彼得堡、柏林和全法国的统计资料,得出几乎完全一致的男婴出生数与女婴出生数的比值约为 22∶21,即在全体出生婴儿中,男婴约占 51.2%,女婴约占 48.8%,或者说男女性别比约为 104.8%.

之后,经过大量统计资料的验证,全世界各种人种的婴儿性别比都遵循这一规律,如果与此标准差异较大,就一定存在某种人为的因素. 如拉普拉斯统计了法国巴黎 1745～1784 年 40 年间的男婴出生率为 50.06%,与男婴标准出生率相差 1.14%. 经过调查,发现当时的巴黎人"重女轻男",有抛弃男婴的陋俗,以致歪曲了出生率的真相. 日本的统计也基本符合这一标准,仅有几个例外,如明治 38 年: 男女性别比为 102.7%,明治 39 年: 108.7%,明治 40 年: 102.7%,原因是明治 39 年是"丙午年",由于迷信,这一年生的女婴,分别假报到明治 38 年和明治 40 年. 这种现象在现在的日本仍然存在,如在昭和 41 年还出现过.

我国 1953 年和 1964 年人口普查时统计的男女性别比分别为 107.6%和 105.5%,基本符合国际公认的 105%到 107%的标准. 可是,从 1982 年第三次全国人口普查起,我国婴儿出生的男女性别比开始出现较为严重的失调: 1982 年为 108.5%,到 1987 年 1%人口抽样调查时已达 110.9%,1990 年第四次人口普查为 111.3%,1995 年 1%人口抽样调查时高达 115.6%,到 2000 年第 5 次人口普查时已是 116.9%. 男女出生性别比的失调已引起人们的关注.

6.3 总体特征数的估计

根据第 6.2 节开头的数据，还可以求出北京地区近年来 7 月 25 日至 8 月 10 日的日最高气温的样本平均值为 34.02，我们可将其作为北京地区近年来 7 月 25 日至 8 月 10 日的日最高气温平均值的估计.

在数学中，通常把能反映总体某种特征的量称为总体特征数.

● 怎样通过抽样的方法，用样本的特征数估计总体的特征数呢？

6.3.1 平均数及其估计

某校高一(1)班同学在老师的布置下，用单摆进行测试，以检验重力加速度. 全班同学两人一组，在相同条件下进行测试，得到下列实验数据(单位：m/s^2)：

9.62	9.54	9.78	9.94	10.01	9.66	9.88	9.68	10.32
9.76	9.45	9.99	9.81	9.56	9.78	9.72	9.93	9.94
9.65	9.79	9.42	9.68	9.70	9.84	9.90		

● 怎样利用这些数据对重力加速度进行估计？

n 个实数 a_1，a_2，a_3，…，a_n 的和简记为 $\sum\limits_{i=1}^{n} a_i$，"$\sum$"读作"西格玛".

我们常用算术平均数 $\frac{1}{n}\sum\limits_{i=1}^{n} a_i$（其中 a_i $(i=1, 2, \cdots, n)$ 为 n 个实验数据）作为重力加速度的"最理想"的近似值，它的依据是什么呢？

处理实验数据的原则是使这个近似值与实验数据之间的离差最小，设这个近似值为 x，那么它与 n 个实验值 a_i $(i=1, 2, \cdots, n)$ 的离差分别为 $x-a_1$，$x-a_2$，$x-a_3$，…，$x-a_n$. 由于上述离差有正有负，故不宜直接相加. 可以考虑将各个离差的绝对值相加，研究

$$|x-a_1|+|x-a_2|+\cdots+|x-a_n|$$

取最小值时 x 的值. 但由于含有绝对值，运算不太方便，所以，考虑离差的平方和，即

$$(x-a_1)^2+(x-a_2)^2+\cdots+(x-a_n)^2,$$

当此和最小时，对应的 x 的值作为近似值. 因为

$$\begin{aligned}&(x-a_1)^2+(x-a_2)^2+\cdots+(x-a_n)^2\\=&nx^2-2(a_1+a_2+\cdots+a_n)x+a_1^2+a_2^2+\cdots+a_n^2,\end{aligned}$$

所以当 $x=\frac{a_1+a_2+\cdots+a_n}{n}$ 时离差的平方和最小，故可用

$\dfrac{a_1+a_2+\cdots+a_n}{n}$ 作为表示这个物理量的理想近似值，称其为这 n 个数据 $a_1, a_2, \cdots, a_n$ 的平均数(average)或均值(mean)，一般记为

$$\bar{a}=\frac{a_1+a_2+\cdots+a_n}{n}.$$

这样，我们可以用计算器求得，由高一(1)班学生的实验数据估计的重力加速度的最佳近似值为 $x=9.774(\mathrm{m/s^2})$.

例 1　某校高一年级的甲、乙两个班级(均为 50 人)的语文测试成绩如下(总分：150 分)，试确定这次考试中，哪个班的语文成绩更好一些.

甲班

112	86	106	84	100	105	98	102	94	107
87	112	94	94	99	90	120	98	95	119
108	100	96	115	111	104	95	108	111	105
104	107	119	107	93	102	98	112	112	99
92	102	93	84	94	94	100	90	84	114

乙班

116	95	109	96	106	98	108	99	110	103
94	98	105	101	115	104	112	101	113	96
108	100	110	98	107	87	108	106	103	97
107	106	111	121	97	107	114	122	101	107
107	111	114	106	104	104	95	111	111	110

分析　我们可用一组数据的平均数衡量这组数据的水平，因此，分别求得甲、乙两个班级的平均分即可.

解　用科学计算器分别求得甲班的平均分为 101.1，乙班的平均分为 105.4，故这次考试乙班成绩要好于甲班.

EXCEL & CALCULATOR

Excel 中函数“AVERAGE()”可直接用于计算给定数据的平均数. 如将乙班的成绩输到工作表中 A1：J5 区域后，在某空白单元格中输入“=AVERAGE(A1：J5)”，即得乙班的平均分为 105.38(如图 6-3-1).

K1　　fx　=AVERAGE(A1:J5)

	A	B	C	D	E	F	G	H	I	J	K
1	116	95	109	96	106	98	108	99	110	103	105.38
2	94	98	105	101	115	104	112	101	113	96	
3	108	100	110	98	107	87	108	106	103	97	
4	107	106	111	121	97	107	114	122	101	107	
5	107	111	114	106	104	104	95	111	111	110	

图 6-3-1

输入数据之前，按 SHIFT CLR 1 = 键清除统计存储器.

用计算器计算平均数的方法：

(1) 选择计算模式，按 MODE 2 键，选择“统计”模式；

(2) 输入数据，按 112 DT 即可输入数据 112，仿此将甲班的语文测试成绩全部输入(按 DT DT 键能输入同样的数据两次，若输入 5 个相同的数据 20，可按 20 SHIFT ; 5 DT)；

(3) 求平均数，按 SHIFT S-VAR 1 = 键，即得甲班的平均分为 101.1.

思考

某工厂有经理 1 人，另有 6 名管理人员，5 名高级技工，10 名工人和 1 名学徒. 现在需要增加一名新工人. 小张前来应聘，经理说：“我公司报酬不错，平均工资每周 300 元.”小张工作几天后找到经理说：“你欺骗了我，我问过其他工人，没有一个工人的周工资超过 200 元. 平均工资怎可能是 300 元呢?”经理拿出如下的工资表说：“你看，平均工资就是 300 元.”

表 6-3-1

人员	经理	管理人员	高级技工	工人	学徒	合计
周工资	2 200	250	220	200	100	
人数	1	6	5	10	1	23
合计	2 200	1 500	1 100	2 000	100	6 900

小张通过计算发现本题中总体平均数恰为

$$(2\,200\times1+250\times6+220\times5+200\times10+100\times1)\div23=300,$$

并没有错.

这个问题中，总体平均数能客观地反映工人的工资水平吗? 为什么?

这个问题中，计算平均数的式子可以写为

$$2\,200\times\frac{1}{23}+250\times\frac{6}{23}+220\times\frac{5}{23}+200\times\frac{10}{23}+100\times\frac{1}{23}.$$

一般地，若取值为 $x_1, x_2, \cdots, x_n$ 的频率分别为 $p_1, p_2, \cdots, p_n$，则其平均数为 $x_1p_1+x_2p_2+\cdots+x_np_n$.

例 2 下面是某校学生日睡眠时间的抽样频率分布表(单位：h)，试估计该校学生的日平均睡眠时间.

表 6-3-2

睡眠时间	人数	频率
[6, 6.5)	5	0.05
[6.5, 7)	17	0.17
[7, 7.5)	33	0.33
[7.5, 8)	37	0.37
[8, 8.5)	6	0.06
[8.5, 9]	2	0.02
合计	100	1

分析　要确定这 100 名学生的平均睡眠时间，就必须计算其总睡眠时间。由于每组中的个体睡眠时间只是一个范围，可以用各组区间的组中值近似地表示。

解法 1　总睡眠时间约为

$$6.25\times5+6.75\times17+7.25\times33+7.75\times37+8.25\times6+8.75\times2=739(\text{h}).$$

故平均睡眠时间约为 7.39 h.

解法 2　求组中值与对应频率之积的和

$$6.25\times0.05+6.75\times0.17+7.25\times0.33+7.75\times0.37+8.25\times0.06+8.75\times0.02=7.39(\text{h}).$$

答　估计该校学生的日平均睡眠时间约为 7.39 h.

例 3　某单位年收入在 10 000 到 15 000、15 000 到 20 000、20 000到 25 000、25 000 到 30 000、30 000 到 35 000、35 000 到 40 000 及 40 000 到 50 000 元之间的职工所占的比分别为 10%，15%，20%，25%，15%，10%和 5%，试估计该单位职工的平均年收入.

分析　上述比就是各组的频率.

解　估计该单位职工的平均年收入为

$$12\,500\times10\%+17\,500\times15\%+22\,500\times20\%+27\,500\times25\%+32\,500\times15\%+37\,500\times10\%+45\,000\times5\%=26\,125(\text{元}).$$

答　估计该单位人均年收入约为 26 125 元.

练 习

1. 根据你已有的数据，估计你校高一学生的身高的平均值.
2. 对第 6.2.2 节练习第 2 题，试估算出这批灯泡的平均使用寿命.
3. 从某校全体高考考生的数学成绩中任意抽取 20 名考生的成绩(单位：分，总分：150 分)为 102，105，131，95，83，121，140，100，97，96，95，121，124，135，106，109，110，101，98，97，试估计该校全体考生数学平均成绩.
4. 某教师出了一份共 3 道题的测试卷，每道题 1 分. 全班得 3 分、2 分、1 分和 0 分的学生所占比例分别为 30%，50%，10%和 10%.
   (1) 若全班共 10 人，则平均分是多少？
   (2) 若全班共 20 人，则平均分是多少？
   (3) 如果该班人数未知，能求出该班的平均分吗？

6.3.2 方差与标准差

有甲、乙两种钢筋，现从中各抽取一个样本(如表 6－3－3)检查它们的抗拉强度(单位：kg/mm^2)，通过计算发现，两个样本的平均数均为 125.

表 6－3－3

甲	110	120	130	125	120	125	135	125	135	125
乙	115	100	125	130	115	125	125	145	125	145

● 哪种钢筋的质量较好？

将甲、乙两个样本数据分别标在数轴上，如图 6－3－2 所示.

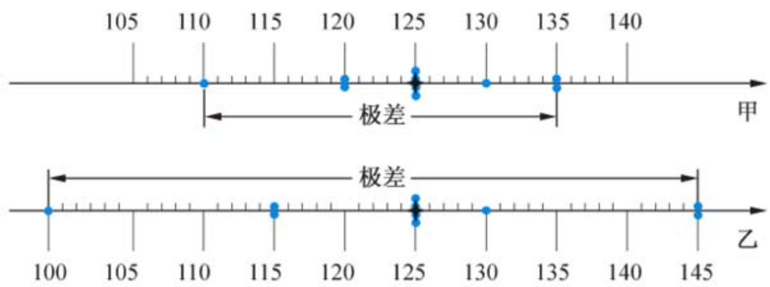


图 6－3－2

从图 6－3－2 中可以看出，乙样本的最小值 100 低于甲样本的最小值 110，最大值 145 高于甲样本的最大值 135，这说明乙种钢筋没有甲种钢筋的抗拉强度稳定.

我们把一组数据的最大值与最小值的差称为极差(range). 从图 6－3－2中可看出，乙的极差较大，数据点较分散；甲的极差小，数据点较集中，这说明甲比乙稳定. 运用极差对两组数据进行比较，操作简单方便，但如果两组数据的集中程度差异不大时，就不容易得出结论.

我们还可以考虑每一抗拉强度与平均抗拉强度的离差，离差越小，稳定性就越高. 结合上节有关离差的讨论，可用各次抗拉强度与平均抗拉强度的差的平方和表示. 由于两组数据的容量可能不同，因此应将上述平方和除以数据的个数，我们把由此所得的值称为这组数据的方差(variance).

因为方差与原始数据的单位不同，且平方后可能夸大了离差的程度，我们将方差开方后的值称为这组数据的标准差(standard deviation). 标准差也可以刻画数据的稳定程度.

一般地，

设一组样本数据 $x_1, x_2, \cdots, x_n$，其平均数为$\overline{x}$，则称

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

为这个样本的方差，其算术平方根 $s=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$ 为样本的标准差，分别简称样本方差、样本标准差.

根据上述方差计算公式可算得甲、乙两个样本的方差分别为 50 和 165，故可认为甲种钢筋的质量好于乙种钢筋.

例 4　甲、乙两种冬水稻试验品种连续 5 年的平均单位面积产量如下(单位：t/hm^2)，试根据这组数据估计哪一种水稻品种的产量比较稳定.

表 6-3-4

品　种	第 1 年	第 2 年	第 3 年	第 4 年	第 5 年
甲	9.8	9.9	10.1	10	10.2
乙	9.4	10.3	10.8	9.7	9.8

解　甲品种的样本平均数为 10，样本方差为

$$\begin{aligned}&[(9.8-10)^2+(9.9-10)^2+(10.1-10)^2+\\&(10-10)^2+(10.2-10)^2]\div 5\\&=0.02,\end{aligned}$$

乙品种的样本平均数也为 10，样本方差为

$$\begin{aligned}&[(9.4-10)^2+(10.3-10)^2+(10.8-10)^2+\\&(9.7-10)^2+(9.8-10)^2]\div 5\\&=0.24,\end{aligned}$$

因为 $0.24>0.02$，

所以，由这组数据可以认为甲种水稻的产量比较稳定.

例 5 为了保护学生的视力，教室内的日光灯在使用一段时间后必须更换. 已知某校使用的 100 只日光灯在必须换掉前的使用天数如下，试估计这种日光灯的平均使用寿命和标准差.

表 6-3-5

天 数	151～180	181～210	211～240	241～270	271～300	301～330	331～360	361～390
灯泡数	1	11	18	20	25	16	7	2

分析 用每一区间内的组中值作为相应日光灯的使用寿命，再求平均寿命.

解 各组中值分别为 165，195，225，255，285，315，345，375，由此算得平均数约为

$$165\times1\%+195\times11\%+225\times18\%+255\times20\%+$$
$$285\times25\%+315\times16\%+345\times7\%+375\times2\%$$
$$=267.9\approx268(\text{天}).$$

将各组中值对于此平均数求方差得

$$\frac{1}{100}\times[1\times(165-268)^2+11\times(195-268)^2+18\times$$
$$(225-268)^2+20\times(255-268)^2+25\times(285-268)^2+$$
$$16\times(315-268)^2+7\times(345-268)^2+2\times(375-268)^2]$$
$$=2\,128.60(\text{天}^2),$$

故标准差约为 $\sqrt{2\,128.6}\approx46(\text{天})$.

答 估计这种日光灯的平均使用寿命约为 268 天，标准差约为 46 天.

EXCEL & CALCULATOR

在 Excel 中，可分别用函数“VARP()”和“STDEVP()”计算方差和标准差. 也可用计算器，在“统计”模式下输入数据，按 SHIFT S-VAR 2 = 键，得标准差，再按 x^2 键即为方差.

练 习

1. 从两个班级各抽 5 名学生测量身高，数据如下(单位：cm)：甲班的为 160，162，159，160，159，乙班的为 180，160，150，150，160. 试估计哪个班级学生身高波动小.
2. 若 k_1，k_2，…，k_8 的方差为 3，则 $2(k_1-3)$，$2(k_2-3)$，…，$2(k_8-3)$ 的方差为__________.
3. 第 6.2.3 节的练习第 2 题的两名运动员中，哪名成绩更稳定？

4. 利用计算器计算下列两组数据的平均数和标准差.

甲	9.9	10.3	9.8	10.1	10.4	10.0	9.8	9.7
乙	10.2	10.0	9.5	10.3	10.5	9.6	9.8	10.1

习题 6.3

感受·理解

1. 某工厂一个月(30天)中的日产值如下：有2天的产值是5.1万元，有3天的产值是5.2万元，有6天的产值是5.3万元，有8天的产值是5.4万元，有7天的产值是5.5万元，有3天的产值是5.6万元，有1天的产值是5.7万元. 试计算该厂这个月的平均日产值.

2. 为了考察某种大麦穗长的分布情况，在一块试验田里抽取了100穗，量得它们的长度如下(单位：cm)：

6.5	6.4	6.7	5.8	5.9	5.9	5.2	4.0	5.4	5.6
5.8	5.5	6.0	6.5	5.1	6.5	5.3	5.9	5.5	5.8
6.2	5.4	5.0	5.0	6.8	6.0	5.0	5.7	6.0	5.5
6.8	6.0	6.3	5.5	5.0	6.8	6.6	6.0	7.0	6.4
6.4	5.8	5.9	5.7	6.8	6.6	6.0	6.4	5.7	7.4
6.0	5.4	6.5	6.0	6.8	5.8	6.3	6.0	6.3	5.6
5.3	6.4	5.7	6.7	6.2	5.6	6.0	6.7	6.7	6.0
5.5	6.2	6.1	5.3	6.2	6.8	6.6	4.7	5.7	5.7
5.8	5.3	7.0	6.0	6.0	5.9	5.4	6.0	5.2	6.0
6.3	5.7	6.8	6.1	4.5	5.6	6.3	6.0	5.8	6.3

请列出频率分布表，并估计该试验田里麦穗的平均长度.

3. 两台机床同时生产一种零件，在10天中，两台机床每天的次品数如下：

甲	1	0	2	0	2	3	0	4	1	2
乙	1	3	2	1	0	2	1	1	0	1

(1) 哪台机床次品数的平均数较小？

(2) 哪台机床生产状况比较稳定？

4. 从1，2，3，4，5，6这6个数中任取2个，求所有这样的两数之积的平均数.

5. 甲、乙两台半自动车床加工同一型号的产品，各生产1 000只产品中次品数分别用 x 和 y 表示. 经过一段时间的观察，发现 x 和 y 的频率分布如下表，问：哪一台车床的产品质量较好？

x	0	1	2	3
P	0.7	0.1	0.1	0.1

y	0	1	2	3
P	0.5	0.3	0.2	0

思考·运用

6. 某制造商生产长度为 6 cm 的金属棒，抽样检查 40 根，测得每根长度(精确到两位小数)，结果如下(单位：cm)：

6.02	6.01	5.94	5.94	5.97	5.96	5.98	6.01	5.98	5.99
6.00	6.03	5.99	5.97	5.98	6.00	6.03	5.95	6.00	6.00
5.92	5.93	5.99	5.99	6.00	5.95	6.00	5.97	5.96	5.97
6.03	6.01	5.98	5.99	6.04	6.00	6.02	5.97	5.96	5.98

(1) 计算上述样本中金属棒的平均长度；

(2) 画出频率分布直方图；

(3) 如果允许制造商对这种棒与 6 cm 的标准有 0.2%的离差，那么抽样检查中合格的金属棒有几根？合格率是多少？

7. 下面是 60 名男生每分钟脉搏跳动的次数：

72	70	66	74	81	70	74	53	57	62	58	92
72	67	62	91	73	64	65	80	78	67	75	80
83	61	72	72	69	70	76	74	65	84	79	80
76	72	68	65	82	79	71	86	77	69	72	56
70	62	76	56	86	63	73	70	75	73	89	64

(1) 作出上述数据的频率分布直方图；

(2) 根据直方图的各组中值估计总体平均数，并将所得结果与实际的总体平均数相比较，计算误差.

探究·拓展

8. 一位研究化肥的科学家将一片土地划分成 100 个 50 m^2 的小块，并在 50 个小块上施用新化肥. 留下 50 个条件大体相当的小块不施新化肥. 施用新化肥的 50 小块土地的小麦产量如下(单位：kg)：

15	29	22	15	3	30	22	16	5	2
22	13	20	25	42	25	20	38	12	29
14	21	26	13	21	27	13	21	11	18
10	18	24	24	36	34	23	18	10	9
17	23	33	8	16	23	31	16	23	40

没有施用新化肥的 50 小块土地上的产量如下(单位：kg)：

23	16	16	17	22	3	10	10	8	14
16	5	24	16	32	23	15	18	9	21
4	24	5	24	15	2	15	25	17	29
33	39	16	17	2	15	17	17	26	13
26	11	18	19	12	20	27	12	28	22

你认为新化肥已经取得成功了吗？

6.4 线性回归方程

在实际问题中，变量之间的常见关系有如下两类：

一类是确定性函数关系，变量之间的关系可以用函数表示. 例如，圆的面积 S 与半径 r 之间就是确定性函数关系，可以用函数 $S=\pi r^2$ 表示.

一类是相关关系(correlation)，变量之间有一定的联系，但不能完全用函数来表达. 例如，人的体重 y 与身高 x 有关. 一般来说，身高越高，体重越重，但不能用一个函数来严格地表示身高与体重之间的关系. 下面问题中热茶销量与气温之间具有相关关系.

某小卖部为了了解热茶销售量与气温之间的关系，随机统计并制作了某 6 天卖出热茶的杯数与当天气温的对比表：

表 6-4-1

气温/℃	26	18	13	10	4	−1
杯　数	20	24	34	38	50	64

如果某天的气温是−5℃，你能根据这些数据预测这天小卖部卖出热茶的杯数吗？

为了了解热茶销量与气温的大致的关系，我们以横坐标 x 表示气温，纵坐标 y 表示热茶销量，建立平面直角坐标系，将表中数据构成的 6 个数对所表示的点在坐标系内标出，得到图 6-4-1. 今后我们称这样的图为散点图(scatter diagram).

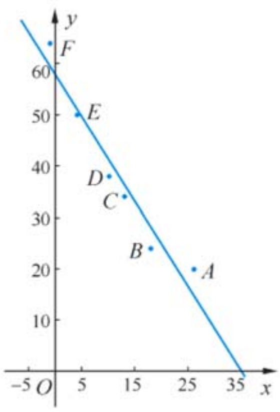


图 6-4-1

从图 6-4-1 可以看出，这些点散布在一条直线的附近，故可用一个线性函数近似地表示热茶销量与气温之间的关系.

● 选择怎样的直线近似地表示热茶销量与气温之间的关系？

我们有多种思考方案：

(1) 选择能反映直线变化的两个点，例如取过(4，50)，(18，24)这两点的直线；

(2) 取一条直线，使得位于该直线一侧和另一侧点的个数基本相同；

(3) 多取几组点，确定几条直线方程，再分别算出各条直线斜率、截距的平均值，作为所求直线的斜率、截距；

……

怎样的直线最好呢？

用方程为 $\hat{y}=bx+a$ 的直线拟合散点图中的点，应使得该直线与散点图中的点最接近. 那么，怎样衡量直线 $\hat{y}=bx+a$ 与图中六个点的接近程度呢？

我们将表中给出的自变量 x 的六个值代入直线方程，得到相应的六个 $\hat{y}$ 值：

$$26b+a,\ 18b+a,\ 13b+a,\ 10b+a,\ 4b+a,\ -b+a.$$

这六个值与表中相应的六个 $\hat{y}$ 的实际值应该越接近越好. 所以，我们用类似于估计总体平均数时的思想，考虑如下离差平方和

$$\begin{aligned}Q(a,b)&=(26b+a-20)^2+(18b+a-24)^2+\\&\quad(13b+a-34)^2+(10b+a-38)^2+\\&\quad(4b+a-50)^2+(-b+a-64)^2\\&=1\,286b^2+6a^2+140ab-3\,820b-460a+10\,172.\end{aligned}$$

$Q(a,b)$ 是直线 $\hat{y}=bx+a$ 与各散点在垂直方向(纵轴方向)上的距离的平方和，可以用来衡量直线 $\hat{y}=bx+a$ 与图中六个点的接近程度. 所以，设法取 a，b 的值，使 $Q(a,b)$ 达到最小值. 这种方法叫做最小二乘法.

先把 a 看做常数，那么 Q 是关于 b 的二次函数. 易知，当 $b=-\dfrac{140a-3\,820}{2\times1\,286}$ 时，Q 取得最小值.

同理，把 b 看做常数，那么 Q 是关于 a 的二次函数. 当 $a=-\dfrac{140b-460}{12}$ 时，Q 取得最小值.

因此，当

$$\begin{cases}b=-\dfrac{140a-3\,820}{2\times1\,286},\\ a=-\dfrac{140b-460}{12}\end{cases}$$

时，Q 取得最小值，由此解得

$$b\approx-1.647\,7,\ a\approx57.556\,8.$$

利用所求线性函数作预测时，通常在自变量 x 的数据范围内有效. 若是自变量 x 超出数据范围作预测(通常称为“外推”)时应慎重.

所求直线方程为

$$\hat{y}=-1.647\,7x+57.556\,8.$$

当 $x=-5$ 时，$\hat{y}\approx66$，故当气温为 -5℃时，热茶销量约为 66 杯.

像这样能用直线方程 $\hat{y}=bx+a$ 近似表示的相关关系叫做**线性相关关系**(linear correlation).

一般地，设有 (x,y) 的 n 对观察数据如下：

表 6-4-2

x	x_1	x_2	x_3	$\cdots$	x_n
y	y_1	y_2	y_3	$\cdots$	y_n

当 a，b 使

$$Q=(y_1-bx_1-a)^2+(y_2-bx_2-a)^2+\cdots+(y_n-bx_n-a)^2$$

取得最小值时，就称 $\hat{y}=bx+a$ 为拟合这 n 对数据的线性回归方程 (linear regression equation)，将该方程所表示的直线称为回归直线.

仿照前面的方法 3，可得线性回归方程 $\hat{y}=bx+a$ 中的系数 a，b 满足：

$$\begin{cases}\left(\sum\limits_{i=1}^{n}x_i^2\right)b+\left(\sum\limits_{i=1}^{n}x_i\right)a=\sum\limits_{i=1}^{n}x_iy_i,\\ \left(\sum\limits_{i=1}^{n}x_i\right)b+na=\sum\limits_{i=1}^{n}y_i,\end{cases}$$

由此二元一次方程组便可依次求出 b，a 的值：

有兴趣的同学可以尝试推导.

$$\begin{cases}b=\dfrac{n\sum\limits_{i=1}^{n}x_iy_i-\left(\sum\limits_{i=1}^{n}x_i\right)\left(\sum\limits_{i=1}^{n}y_i\right)}{n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2},\\ a=\overline{y}-b\overline{x}.\end{cases}\quad\cdots\cdots\cdots\cdots\cdots\ (*)$$

例 1　下表为某地近几年机动车辆数与交通事故数的统计资料，请判断机动车辆数与交通事故数之间是否具有线性相关关系，如果具有线性相关关系，求出线性回归方程；如果不具有线性相关关系，说明理由.

表 6-4-3

机动车辆数 x/千台	95	110	112	120	129	135	150	180
交通事故数 y/千件	6.2	7.5	7.7	8.5	8.7	9.8	10.2	13

用计算器计算 $\sum\limits_{i=1}^{n}x_i$，$\sum\limits_{i=1}^{n}x_i^2$ 可一次完成：先用“M+”键输入 x_i $(i=1,2,3,\cdots,n)$，再分别按 $\sum x$，$\sum x^2$ 键可得.

求 $\sum\limits_{i=1}^{n}x_iy_i$ 时，每次将两数之积输入“M+”，最后按 $\sum$ 即可，不必具体写出 x_iy_i 的值.

解　在直角坐标系中描出数据的散点图，直观判断散点在一条直线附近，故具有线性相关关系. 计算相应的数据之和：

$$\sum_{i=1}^{8}x_i=1\,031,\ \sum_{i=1}^{8}y_i=71.6,$$

$$\sum_{i=1}^{8}x_i^2=137\,835,\ \sum_{i=1}^{8}x_iy_i=9\,611.7,$$

将它们代入(∗)式计算得

$$b \approx 0.0774,\ a = -1.0241,$$

所以,所求线性回归方程为

$$\hat{y} = 0.0774x - 1.0241.$$

EXCEL & CALCULATOR

求线性回归方程可借助计算器或计算机来完成. 用 Excel 进行数据拟合,得到图 6-4-2.

上面得到的直线即为回归直线,显示的 R^2 的值越接近于 1,其拟合效果越好.

用 Excel 和计算器进行数据拟合,请参阅数学 1(必修)第 2.6 节中的内容.

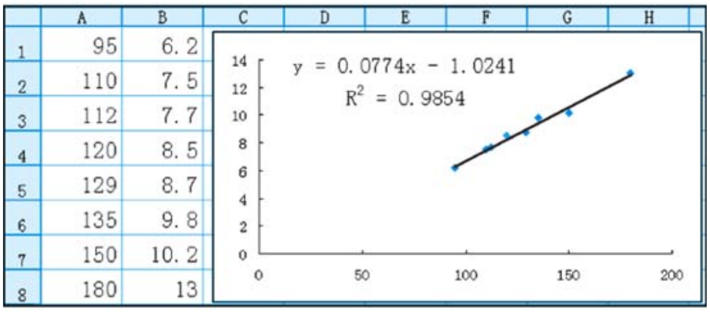


图 6-4-2

用计算器求回归直线的方法(CASIO fx-82MS)如下.

(1) 按 $\boxed{\text{MODE}}$ 键,选择“回归计算”模式(REG);

(2) 进入 REG 模式后,按 $\boxed{1}$ 选择回归种类“线性(Lin)”;

(3) 输入数据,按键 95 $\boxed{,}$ 6.2 $\boxed{\text{DT}}$,即可输入一组数据(95, 6.2),将 8 组数据仿此全部输入;

(4) 计算回归直线方程 $\hat{y} = bx + a$ 中的系数,

按 $\boxed{\text{SHIFT}}\boxed{\text{S-VAR}}\boxed{▶}\boxed{▶}\boxed{1}\boxed{=}$ 得回归系数

$$a = -1.024111886,$$

按 $\boxed{\text{SHIFT}}\boxed{\text{S-VAR}}\boxed{▶}\boxed{▶}\boxed{2}\boxed{=}$ 得回归系数

$$b = 0.07739369;$$

(5) 如果要预测机动车辆数为 200 千台时的交通事故数 y(千件)的值,即计算当 $x = 200$ 时 y 的值,可直接

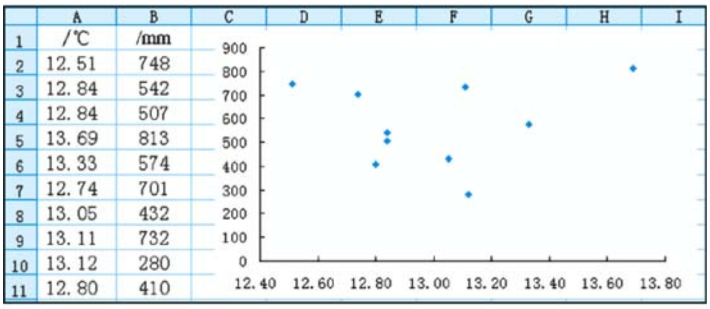
按 200 $\boxed{\text{SHIFT}}\boxed{\text{S-VAR}}\boxed{▶}\boxed{▶}\boxed{▶}\boxed{2}\boxed{=}$ 得

$$\hat{y} = 14.45462625.$$

共有线性(Lin)、对数(Log)、指数(Exp)、乘方(Pwr)、逆(Inv)和二次(Quad)六种回归模型可供选择.

思 考

图 6-4-3 是 1991 年至 2000 年北京地区年平均气温(单位:℃)与年降雨量(单位:mm)的散点图. 此时求得的回归直线方程有意义吗?



	A	B
1	/℃	/mm
2	12.51	748
3	12.84	542
4	12.84	507
5	13.69	813
6	13.33	574
7	12.74	701
8	13.05	432
9	13.11	732
10	13.12	280
11	12.80	410

图 6－4－3

一般地，用回归直线进行拟合的一般步骤为：

(1) 作出散点图，判断散点是否在一条直线附近；

(2) 如果散点在一条直线附近，用公式(＊)求出 a，b，并写出线性回归方程.

练　习

1. 下面是我国居民生活污水排放量的一组数据(单位：10^8 t)，试分别估计 1996 年和 2004 年我国居民生活污水排放量.

年　份	1995	1996	1997	1998
排放量	151		189.1	194.8
年　份	1999	2000	2001	2002
排放量	203.8	220.9	227.7	232.3

2. 一个工厂在某年里每月产品的总成本 y(单位：万元)与月产量 x(单位：万件)之间有如下一组数据：

x	1.08	1.12	1.19	1.28	1.36	1.48	1.59	1.68	1.80	1.87	1.98	2.07
y	2.25	2.37	2.40	2.55	2.64	2.75	2.92	3.03	3.14	3.26	3.36	3.50

(1) 画出散点图；

(2) 求出月总成本 $\hat{y}$ 与月产量 x 之间的线性回归方程.

习题 6.4

感受·理解

1. 某种产品的广告费支出 x 与销售额 y(单位：百万元)之间有如下对应数据：

x	2	4	5	6	8
y	30	40	60	50	70

(1) 画出散点图；

(2) 求线性回归方程.

2. 每立方米混凝土的水泥用量 x(单位：kg)与 28 天后混凝土的抗压强度(单位：kg/cm^2)之间的关系有如下对应数据：

x	150	160	170	180	190	200
y	56.9	58.3	61.1	64.6	68.1	71.3
x	210	220	230	240	250	260
y	74.1	77.4	80.2	82.6	86.4	89.7

(1) 画出散点图；

(2) 求线性回归方程.

3. 在国民经济中，社会生产和货运之间有着密切的关系，下面列出 1991 到 2000 年某地区货运量与工业总产值的统计资料：

年　份	1991	1992	1993	1994	1995
总产值 $x/10^9$ 元	2.8	2.9	3.2	3.2	3.4
货运量 $y/10^4$ t	25	27	29	32	34
年　份	1996	1997	1998	1999	2000
总产值 $x/10^9$ 元	3.2	3.3	3.7	3.9	4.2
货运量 $y/10^4$ t	36	35	39	42	45

(1) 画出散点图；

(2) 求出变量 x，y 之间的线性回归方程.

思考·运用

4. 试根据第 6.1.2 节练习第 3 题所得数据，分析你班学生的两臂平展的长度与身高是否具有线性相关关系. 若存在线性相关关系，求出线性回归方程；若不存在，请说明理由.

链　接

相 关 系 数

观察如图 6－4－4 所示的第 6.4 节节首问题和例 1 中数据的散点图：

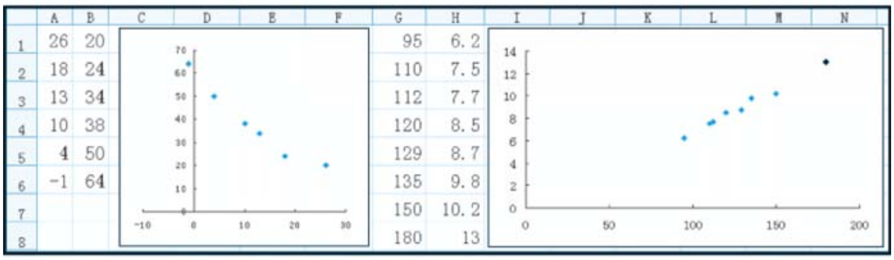


图 6－4－4

它们的线性回归方程分别为

$$\hat{y}=-1.6477x+57.5568,$$

$$\hat{y}=0.0774x-1.0241,$$

两者有何异同？

前者随着自变量的增大，函数值在减小，对应的一次项系数为负

数；后者随着自变量的增大，函数值在增大，对应的一次项系数为正数.

一般地，在线性回归方程 $\hat{y}=bx+a$ 中，当一次项系数 b 为正数时，我们就称这两个变量正相关；当 b 为负数时，就称这两个变量负相关.

另一方面，两个散点图中的点与回归直线的偏离程度也不相同，说明两个变量之间的相关程度也有区别. 对于变量 y 与 x 的一组观测值来说，我们常用 r 来描述线性相关的程度：

$$r=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}},$$

即

$$r=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right)\left(\sum_{i=1}^{n}y_i^2-n\overline{y}^2\right)}}.$$

这个数值 r 称为 $\hat{y}$ 与 x 的样本相关系数，简称相关系数.

当 $r>0$ 时，$\hat{y}$ 与 x 正相关；当 $r<0$ 时，$\hat{y}$ 与 x 负相关. 可以证明 $|r|\leqslant 1$. $|r|$ 越接近 1，线性相关程度越高；$|r|$ 越接近于 0，线性相关程度越低.

EXCEL & CALCULATOR

Excel 提供了统计函数“CORREL()”，用于计算两个变量间的相关系数. 下面是第 6.4 节的节首问题和例 1 中相关系数的计算方法：

将节首问题中的数据输入工作表 A1：B6 区域(图 6－4－4)，在空白单元格中输入“＝CORREL(A1：A6，B1：B6)”，即得“气温”与“杯数”间的相关系数为－0.97.

将例 1 中的数据输入工作表 G1：H8 区域(图 6－4－4)，在空白单元格中输入“＝CORREL(G1：G8，H1：H8)”，可知机动车辆数和交通事故数间的相关系数为 0.993.

例如，某公司购进一新型设备，为了分配合适的工人操纵设备，需进行该设备的工人劳动生产率与工龄之间的相关分析. 下表是 12 个工人操纵新设备的劳动生产率的试验记录：

表 6－4－4

工龄 x/a		5	5	6	6	6	7	7	8	8	9	10	10
生产率	min/件	8.5	8.3	8.0	8.0	7.8	7.2	7.0	6.5	6.5	6.0	6.2	6.0
	件 y/h	7.1	7.2	7.5	7.5	7.7	8.3	8.6	9.2	9.2	10	9.7	10

请根据上表计算工龄与劳动生产率的相关系数，并进行相关分析.

解 将数据输到 Excel 工作表中，在单元格 A4 中输入“=CORREL(B1：M1，B3：M3)”，得 $r_1\approx0.971$（图 6－4－5），说明工龄与劳动生产率之间存在高度正相关关系.

A4 　 =CORREL(B1:M1,B3:M3)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	工龄	5	5	6	6	6	7	7	8	8	9	10	10
2	分/件	8.5	8.3	8	8	7.8	7.2	7	6.5	6.5	6	6.2	6
3	件/小时	7.1	7.2	7.5	7.5	7.7	8.3	8.6	9.2	9.2	10	9.7	10
4	0.97133334												
5	-0.96280146												

图 6－4－5

如果劳动生产率以每件所花时间（单位：min）计算，工龄与劳动生产率的相关系数由“＝CORREL（B1：M1，B2：M2）”得 $r_2\approx-0.963$. 说明工龄长短与劳动生产率具有高度的负相关关系. 工龄越长，每件产品耗用的工时（单位：min）越少，说明经验和知识积累对于提高劳动生产率有着非常重要的意义.

用计算器求相关系数的方法如下：

恢复计算器的默认设置，可按 SHIFT CLR 2 =

（1）选择计算模式，按 MODE 3 1 键，选择“回归计算”模式，类型为“线性”；

（2）输入数据，按 5 , 8.5 DT，可输入一组数据（5，8.5），仿此输入全部数据；

（3）计算相关系数，按 SHIFT S-VAR ▶ ▶ 3 = 键可得相关系数.

实习作业

1. 对你校高一年级 15 周岁的男生（或女生）身高进行调查，把抽取的 60 名男生（或女生）测得的身高作为样本，将所得数据进行分析处理，完成相应的实习报告.

2. 自己（或分组）选择适当的课题，进行统计研究，并写出报告 .

下列课题供参考：

（1）在本校范围内就学生的视力情况进行抽样调查，对数据进行统计分析，写出实习报告，并就提高学生视力提出建设性的意见.

（2）在本校范围内就学生的开支情况（学费、生活费、资料费、交通费、活动费、零花钱等）进行抽样调查，对数据进行统计分析，写出实习报告，并就学生开支情况提出你的看法.

（3）对你校学生分年龄、男女，进行身高抽样调查，就男、女生身高及其增长速度等方面进行统计分析，设计调查表，并写出实习报告.

附：实习报告表

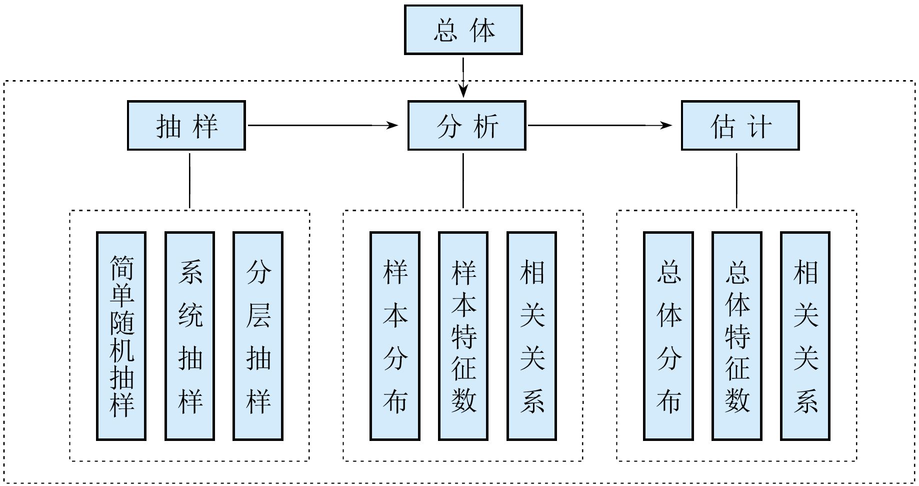
实 习 报 告

______年______月______日

题目	
基本要求	
抽样说明	
样本数据	
频率分布表	
频率分布直方图	
样　本平均数	
结果分析	
负责人及参加人　员	
指导教师审核意　见	
备　注	

本章回顾

本章介绍了从总体中抽取样本的常用方法，并通过实例，研究了如何利用样本对总体的分布规律、整体水平、稳定程度及相关关系等特性进行估计和预测.



当总体容量大或检测具有一定的破坏性时，可以从总体中抽取适当的样本，通过对样本的分析、研究，得到对总体的估计，这就是统计分析的基本过程. 而用样本估计总体就是统计思想的本质.

要准确估计总体，必须合理地选择样本，我们学习的是最常用的三种抽样方法. 获取样本数据后，将其用频率分布表、频率直方图、频率折线图或茎叶图表示后，蕴含于数据之中的规律得到直观的揭示. 运用样本的平均数可以对总体水平作出估计，用样本的极差、方差（标准差）可以估计总体的稳定程度.

对两个变量的样本数据进行相关性分析，可发现存在于现实世界中的回归现象. 用最小二乘法研究回归现象，得到的线性回归方程可用于预测和估计，为决策提供依据.

总之，统计的基本思想是从样本数据中发现统计规律，实现对总体的估计.

复习题

感受·理解

1. 将容量为 100 的样本数据分为 8 个组,如下表:

组　号	1	2	3	4	5	6	7	8
频　数	10	13		14	15	13	12	9

则第 3 组的频率为(　　).

A. 0.14　　B. 0.03　　C. 0.07　　D. 0.21

2. 调查某班学生的平均身高,从 50 位学生中抽取$\frac{1}{10}$,应如何抽样?如果知道男女生的身高显著不同(男生 30 人,女生 20 人),又如何抽样?

3. 140 位选手参加高尔夫球赛,他们的成绩统计如下:

杆　数	65	66	67	68	69	70	71	72	73	74	75
选手数	1	2	5	3	8	17	20	31	22	21	10

(1) 列出频率分布表;

(2) 作出频率分布直方图.

4. 对你班学生进行抽样调查,了解零花钱的使用情况.

5. 200 名学生参与研究性学习,每人仅参加一个课题组. 其中参加文学类的有 33 人,参加理化类的有 30 人,参加数学类的有 62 人,参加社会科学类的有 47 人,参加信息类的有 28 人.

(1) 列出学生参加各类课题组的频率分布表;

(2) 画出表示频率分布的直方图.

6. 对某种电子元件进行寿命追踪调查,情况如下表:

寿命/h	100～200	200～300	300～400	400～500	500～600
个　数	20	30	80	40	30

(1) 列出频率分布表;

(2) 画出频率分布直方图;

(3) 估计该电子元件寿命在 100 h～400 h 内所占的比;

(4) 估计该电子元件寿命在 400 h 以上所占的比.

7. 甲、乙两名学生某门课程的 5 次测试成绩依次为 60, 80, 70, 90, 70 和 80, 65, 70, 80, 75,因为________________,所以学生________成绩稳定.

8. 下面是从某校高一学生中抽取的 20 名学生的学习用书的重量(单位: kg):

8.4	10.1	6.3	7.1	6.2	6.5	7.6	8.0	8.5	6.4
10.3	8.8	5.2	4.6	7.8	3.9	4.8	7.2	8.0	6.8

(1) 作出频率分布直方图；

(2) 利用频率分布直方图的组中值对总体平均数及方差进行估计，并与实际结果进行比较.

9. 下面是南京市与哈尔滨市 2001 年 12 个月的月平均气温(单位：℃)，试分析这两个城市的月平均气温是否具有线性相关关系. 若具有，求出线性回归方程；若不具有，说明理由.

月　份	1	2	3	4	5	6
南京月平均气温	2	3.8	8.4	14.8	19.9	24.5
哈尔滨月平均气温	−19.4	−15.4	−4.8	6	14.3	20
月　份	7	8	9	10	11	12
南京月平均气温	28	27.8	22.7	16.9	10.5	4.4
哈尔滨月平均气温	22.8	21.1	14.4	5.6	−5.7	−15.6

10. 下面是某班学生的父母年龄的茎叶图，试比较这些同学的父亲与母亲的平均年龄.

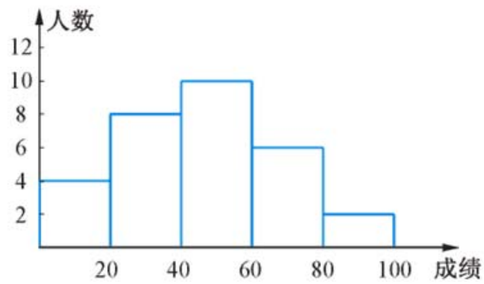
父亲		母亲
88	3	56899
9988865432110	4	023344467899
8775421	5	122357
1	6	

(第 10 题)

思考·运用

11. 抽样统计你所在的学校学生的身高与心率，并用散点图表示数据，判断两者是否线性相关. 若是，求出线性回归方程；若不是，请说明理由.

12. 下面是一次考试结果的频数分布直方图，请据此估计这次考试的平均分.



(第 12 题)

13. 请你搜集有关的数据，估算一下我国 2004 年 18 岁的人口数.

探究·拓展

14. 青年歌手大奖赛有 10 名选手参加，12 名评委给出的评判分数如下表：

歌　手	1	2	3	4	5	6	7	8	9	10
评委 1	9.07	8.98	8.80	8.80	9.00	8.81	9.24	8.91	8.98	9.05
评委 2	9.12	8.92	8.83	8.86	9.05	8.85	9.28	9.02	9.06	9.10
评委 3	9.15	8.95	8.91	8.90	9.12	9.00	9.28	9.05	9.10	9.12
评委 4	9.15	9.00	9.05	8.93	9.15	9.00	9.29	9.15	9.10	9.18
评委 5	9.15	9.01	9.10	8.93	9.15	9.04	9.30	9.16	9.10	9.20
评委 6	9.18	9.02	9.17	9.03	9.15	9.15	9.32	9.20	9.10	9.20
评委 7	9.21	9.15	9.20	9.05	9.16	9.15	9.35	9.21	9.14	9.20
评委 8	9.21	9.20	9.20	9.05	9.17	9.18	9.35	9.21	9.17	9.24
评委 9	9.24	9.20	9.24	9.15	9.18	9.19	9.38	9.21	9.17	9.29
评委 10	9.26	9.23	9.28	9.17	9.18	9.24	9.40	9.28	9.22	9.30
评委 11	9.30	9.30	9.29	9.21	9.24	9.25	9.45	9.29	9.29	9.30
评委 12	9.35	9.32	9.31	9.21	9.31	9.32	9.45	9.30	9.40	9.35

(1) 确定歌手的名次；

(2) 对评委的评判水平给出评价，以便确定下次聘请的 10 名评委.

概率论的诞生，虽然渊源于靠碰运气取胜的游戏，但在今天，却已成为人类知识的最重要的一部分.

——拉普拉斯

足球比赛用抛掷硬币的方式决定场地，这是否公平？

某班的 50 名学生中，有两名学生的生日相同的可能性有多大？

路口有一红绿灯，东西方向的红灯时间为 45 s，绿灯时间为60 s. 从东向西行驶的一辆汽车通过该路口，遇到红灯的可能性有多大？

……

生活中存在大量需要确定“可能性”大小的事件. 概率论就是研究可能性大小的数学分支，它探讨随机现象的规律性，为人们认识世界提供了重要的模式和方法.

- 如何数学地刻画可能性问题？
- 怎样确定一件事发生的可能性大小？

7.1 随机事件及其概率

7.1.1 随机现象

观察下列现象：

(1) 在标准大气压下水加热到 100℃，沸腾；

(2) 导体通电，发热；

(3) 同性电荷，互相吸引；

(4) 实心铁块丢入水中，铁块浮起；

(5) 买一张福利彩票，中奖；

(6) 掷一枚硬币，正面向上.

● 这些现象各有什么特点？

(1)、(2)两种现象必然发生，(3)、(4)两种现象不可能发生，(5)、(6)两种现象可能发生，也可能不发生.

试举出一些确定性现象和随机现象的实例.

在一定条件下，事先就能断定发生或不发生某种结果，这种现象就是确定性现象. 在一定条件下，某种现象可能发生，也可能不发生，事先不能断定出现哪种结果，这种现象就是随机现象. 在自然界和人类社会的生产与生活中，存在着大量的确定性现象和随机现象.

对于某个现象，如果能让其条件实现一次，就是进行了一次试验. 而试验的每一种可能的结果，都是一个事件.

我们看到，如果把(1)、(2)的条件各实现一次，则一定出现“沸腾”与“发热”的结果，“沸腾”与“发热”都是一个事件. 这种在一定的条件下，必然会发生的事件叫做必然事件(certain event).

当(3)、(4)的条件各实现一次，则“吸引”与“浮起”也都是一个事件，但这两个事件都是不可能发生的. 在一定条件下，肯定不会发生的事件叫做不可能事件(impossible event).

当(5)、(6)的条件各实现一次时，“中奖”及“正面向上”也都是一个事件，但这两个事件有可能发生，也可能不发生. 在一定条件下，可能发生也可能不发生的事件叫做随机事件(random event).

必然事件与不可能事件反映的就是在一定条件下的确定性现象，而随机事件反映的则是随机现象.

以后我们用 A，B，C 等大写英文字母表示随机事件，简称为事件.

例 1 试判断下列事件是随机事件、必然事件还是不可能事件：

(1) 我国东南沿海某地明年将 3 次受到热带气旋的侵袭；

(2) 若 a 为实数，则 $|a| \geqslant 0$；

(3) 某人开车通过 10 个路口都将遇到绿灯；

(4) 抛一石块，下落；

(5) 一个正六面体的六个面分别写有数字 1，2，3，4，5，6，将它抛掷两次，向上的面的数字之和大于 12.

解 由题意知，(2)、(4)为必然事件；(5)是不可能事件；(1)、(3)是随机事件.

思考

抛掷一枚硬币，正面向上是一个随机事件. 请与同桌一起进行 100 次试验，统计正面向上的频数及频率.

正面向上的频率是否具有某种规律？

对于随机现象，虽然知道会出现哪些结果，却事先不能确定具体会发生哪一个结果，即无法确定某个随机事件是否发生. 但是，如果在相同条件下大量重复试验时，可以发现随机事件的发生与否呈现出某种规律性. 概率论正是研究随机现象这种数量规律的一个数学分支.

练习

1. 有下列事件：① 连续掷一枚硬币两次，两次都出现正面朝上；② 异性电荷，相互吸引；③ 在标准大气压下，水在 1℃结冰. 其中是随机事件的有(　　).

A. ②　　B. ③　　C. ①　　D. ①、③

2. 指出下列事件是必然事件、不可能事件还是随机事件：

(1) 在一条公路上，交警记录某一小时通过汽车超过 500 辆；

(2) 若 a 为实数，则 $|a+1|+|a+2|=0$；

(3) 北京地区每年 1 月份月平均气温低于 7 月份的月平均气温；

(4) 在常温常压下，石墨能变成金刚石；

(5) 发射一枚炮弹，命中目标；

(6) 明天下雨.

3. 给出下列事件：① 明天进行的某场足球赛的比分是 3∶1；② 下周一某地的最高气温与最低气温相差 10℃；③ 同时掷两颗骰子，向上一面的两个点数之和不小于 2；④ 射击 1 次，命中靶心；⑤ 当 x 为实数时，$x^2+4x+4<0$. 其中，必然事件有________，不可能事件有________，随机事件有________.

4. 请举出一些必然事件、不可能事件和随机事件的实例.

7.1.2 随机事件的概率

我们已经学习了用概率表示一个事件在一次试验或观测中发生的可能性的大小，它是 0～1 之间的一个数. 将这个事件记为 A，用 $P(A)$表示事件 A 发生的概率. 对于任意一个随机事件 A，$P(A)$必须满足如下基本要求：

$$0 \leqslant P(A) \leqslant 1.$$

● 怎样确定一事件发生的概率呢？

奥地利遗传学家孟德尔(G. Mendel，1822～1884)用豌豆进行杂交试验，下表为试验结果(其中 F_1 为第一子代，F_2 为第二子代)：

表 7-1-1

性　状	F_1 的表现	F_2 的表现		
种子的形状	全部圆粒	圆粒　5 474	皱粒　1 850	圆粒∶皱粒≈2.96∶1
茎 的 高 度	全部高茎	高茎　787	矮茎　277	高茎∶矮茎≈2.84∶1
子叶的颜色	全部黄色	黄色　6 022	绿色　2 001	黄色∶绿色≈3.01∶1
豆荚的形状	全部饱满	饱满　882	不饱满　299	饱满∶不饱满≈2.95∶1

孟德尔发现第一子代对于一种性状为必然事件，其可能性为 100%，另一性状的可能性为 0，而第二子代对于前一性状的可能性约为 75%，后一性状的可能性约为 25%. 通过进一步研究，他发现了生物遗传的基本规律.

实际上，孟德尔是从某种性状发生的频率作出估计的.

在《算法初步》一章中，我们曾设计了一个抛掷硬币的模拟试验. 图 7-1-1 是连续 8 次模拟试验的结果：

在 http://www.1088.com.cn/move/003/下载课件.

	A	B
1	模拟次数 10	正面向上的频率0.3
2	模拟次数 100	正面向上的频率0.53
3	模拟次数 1000	正面向上的频率0.52
4	模拟次数 5000	正面向上的频率0.4996
5	模拟次数 10000	正面向上的频率0.506
6	模拟次数 50000	正面向上的频率0.50118
7	模拟次数 100000	正面向上的频率0.49904
8	模拟次数 500000	正面向上的频率0.50019

图 7-1-1

在第 7.1.1 节中所做的抛掷硬币试验中的结论能否说明这一点？

我们看到，当模拟次数很大时，正面向上的频率值接近于常数 0.5，并在其附近摆动.

再看表 7－1－2 和表 7－1－3.

表 7－1－2　π 的前 n 位小数中数字 6 出现的频率

n	数字 6 出现次数	数字 6 出现频率
100	9	0.090 000
200	16	0.080 000
500	48	0.096 000
1 000	94	0.094 000
2 000	200	0.100 000
5 000	512	0.102 400
10 000	1 004	0.100 400
50 000	5 017	0.100 340
1 000 000	99 548	0.099 548

表 7－1－3　鞋厂某种成品鞋质量检验结果

抽取产品数 n	20	50	100	200	500	1 000
优等品数 m	18	48	96	193	473	952
优等品频率 $\frac{m}{n}$	0.9	0.96	0.96	0.965	0.946	0.952

从表 7－1－2 可以看出：数字 6 在 π 的各位小数数字中出现的频率值接近于常数 0.1，并在其附近摆动. 如果统计 0 至 9 这 10 个数字在 π 的各位数字中出现的频率值，可以发现它们都是接近于常数 0.1，并在其附近摆动.

从表 7－1－3 可以看出，当抽取的样品数很多时，优等品的频率接近于常数 0.95，并在它的附近摆动.

从以上几个实例可以看出：在相同条件下，随着试验的次数的增加，随机事件发生的频率会在某个常数附近摆动并趋于稳定，我们可以用这个常数来刻画该随机事件发生的可能性大小，而将频率作为其近似值.

一般地，如果随机事件 A 在 n 次试验中发生了 m 次，当试验的次数 n 很大时，我们可以将事件 A 发生的频率 $\frac{m}{n}$ 作为事件 A 发生的概率的近似值，即

$$P(A) \approx \frac{m}{n}.$$

所以,在表7-1-2所示的实例中,我们用0.1作为所考虑事件的概率,而在表7-1-3所示的实例中,我们用0.95作为相应事件的概率.

对必然事件和不可能事件,我们将它们作为随机事件的特例考虑,分别用Ω和$\varnothing$表示,显然

$$P(\Omega)=1,\ P(\varnothing)=0.$$

这是概率必须满足的第二个基本条件.

例2　某市统计近几年新生儿出生数及其中男婴数(单位:人)如下:

表7-1-4

时　间	1999年	2000年	2001年	2002年
出生婴儿数	21 840	23 070	20 094	19 982
出生男婴数	11 453	12 031	10 297	10 242

(1) 试计算男婴各年出生频率(精确到0.001);

(2) 该市男婴出生的概率约是多少?

解　(1) 1999年男婴出生的频率为

$$\frac{11\,453}{21\,840}\approx 0.524.$$

同理可求得2000年、2001年和2002年男婴出生的频率分别为0.521,0.512,0.512;

(2) 各年男婴出生的频率在0.51~0.53之间,故该市男婴出生的概率约为0.52.

练　习

1. 每道选择题有4个选择支,其中只有1个选择支是正确的.某次考试共有12道选择题,某人说:"每个选择支正确的概率是$\frac{1}{4}$,我每题都选择第一个选择支,则一定有3题选择结果正确."这句话对吗?
2. 掷一枚硬币,连续出现5次正面向上.张欣认为下次出现反面向上的概率大于$\frac{1}{2}$,你同意吗?为什么?
3. 某医院治疗一种疾病的治愈率为10%,那么,前9个病人都没有治愈,第10个人就一定能治愈吗?

习题7.1

感受·理解

1. 指出下列事件中,哪些是随机事件、必然事件或不可能事件:

(1) 任取三条线段,这三条线段恰能组成直角三角形;

(2) 任取一个正方体的三个顶点,这三个顶点不共面;

(3) 从一个三角形的三个顶点各任画一条射线,这三条射线交于一点;
(4) 把 9 写成两个数的和,其中一定有一个数小于 5;
(5) 实数 a, b 不都为 0,但 $a^2+b^2=0$;
(6) 汽车排放尾气,污染环境;
(7) 明天早晨有雾;
(8) 明年 7 月 28 日的最高气温高于今年 8 月 10 日的最高气温.

2. 某城市的天气预报中,有"降水概率预报",例如预报"明天降水概率为 90%",这是指(　　).

A. 明天该地区约 90%的地方会降水,其余地方不降水
B. 明天该地区约 90%的时间会降水,其余时间不降水
C. 气象台的专家中,有 90%认为明天会降水,其余的专家认为不降水
D. 明天该地区降水的可能性为 90%

3. 某射击运动员进行双向飞碟射击训练,各次训练的成绩记录如下:

射击次数	100	120	150	100	150	160	150
击中飞碟数	81	95	123	82	119	127	121
击中飞碟的频率							

(1) 将各次记录击中飞碟的频率填入表中;
(2) 这个运动员击中飞碟的概率约为多少?

4. 对本书附录一中的"随机数表"的前 20 行,统计数字 0 出现的频率,并对随机数表中各个数字出现的概率作出估计.

思考·运用

5. 将一颗骰子掷 600 次,请你估计掷出的点数大于 2 的次数大约是多少.

6. (操作题) 全班学生每人抛掷一次性纸杯 20 次,先分别统计杯口朝下的频数和频率,再分组统计杯口朝下的频数和频率,最后对全班统计杯口朝下的频数和频率. 由此对杯口朝下的概率作出估计.

探究·拓展

7. 在柯南道尔的侦探小说《跳舞的小人》中,福尔摩斯根据英语中字母 e 的使用频率最高,破译了用跳舞人形所写的密码. 在美国作家爱伦·坡的小说《金甲虫》中也有类似的情节. 从网上找若干篇英文文章,用计算机统计字母 a, b, c,…,z 出现的频率,由此估计这 26 个字母在英文文章中出现的概率.

阅　读

尚克斯算错了吗?

英国数学家尚克斯(Shanks, 1812~1882)精于数值计算,尤以计算圆周率 π 的值著名. 他在 1853 年利用公式

$$\frac{\pi}{4}=\arctan\frac{1}{5}-\arctan\frac{1}{239}$$

将 π 值计算到 608 位小数,远超历史上的一切记录;1873 年又将 π 值计算到 707 位小数. 像这样数十年如一日的精神令人赞叹. 1937 年的巴黎博览会上,还把这 707 位小数刻在"发现馆"的天井院内,供人瞻仰.

1940 年，一位英国大学生弗格森（Ferguson）出于好奇，对尚克斯 1853 年算得的 π 值的 608 个数字作了统计，结果得出每个不同数字出现的次数如下表：

表 7－1－5

数　字	0	1	2	3	4	5	6	7	8	9
出现次数	60	62	67	68	64	56	62	44	58	67

这一结果出乎他的意料. 他原来猜想这 10 个数字出现的次数应大致相等. 数字 7 怎么会这样少呢？他想：上帝总不会对 7 怀有歧视吧！尚克斯的计算是否有误？于是他用当时最好的公式和最先进的计算工具（当时电子计算机尚未问世），花了一年时间算出 π 的 710 位小数. 结果发现尚克斯的 π 值从第 528 位上开始出现错误. 后来他和美国人伦奇（Wrech）合作，于 1948 年 1 月共同发表了 π 的 808 位小数值. 再次统计的结果，表明数字 7 出现的次数并不明显少于其他数字. 只是这 808 位数字还嫌少一点.

1974 年，法国学者让・盖尤和芳旦娜小姐合作，对用电子计算机算得的 π 的前 100 万位小数进行统计，结果如下表：

表 7－1－6

数　字	出现次数	频　率	与 0.1 的误差
0	99 959	0.099 959	－0.000 041
1	99 758	0.099 758	－0.000 242
2	100 026	0.100 026	＋0.000 026
3	100 229	0.100 229	＋0.000 229
4	100 230	0.100 230	＋0.000 230
5	100 359	0.100 359	＋0.000 359
6	99 548	0.099 548	－0.000 452
7	99 800	0.099 800	－0.000 200
8	99 985	0.099 985	－0.000 015
9	100 106	0.100 106	＋0.000 106

这是一个惊人的结果！它证实了弗格森的猜想，即 0，1，2，…，9 这 10 个数字在 π 的小数中出现的概率相等，都是 0.1. 不仅如此，有人对无理数 e 的小数作过统计，也发现了类似的规律.

7.2 古典概型

有红心 1，2，3 和黑桃 4，5 这 5 张扑克牌，将其牌点向下置于桌上，现从中任意抽取一张，那么抽到的牌为红心的概率有多大？

若进行大量重复试验，用“出现红心”这一事件的频率估计概率，工作量较大且不够准确.

● 有更好的解决方法吗？

把“抽到红心”记为事件 B，那么事件 B 相当于“抽到红心 1”、“抽到红心 2”、“抽到红心 3”这 3 种情况，而“抽到黑桃”相当于“抽到黑桃 4”、“抽到黑桃 5”这 2 种情况，由于是任意抽取的，可以认为出现这 5 种情况的可能性都相等.

当出现抽到红心 1，2，3 这 3 种情形之一时，事件 B 就发生，于是 $P(B)=\frac{3}{5}$.

在一次试验中可能出现的每一个基本结果称为基本事件(elementary event). 如在上面的问题中，“抽到红心 1”即为一个基本事件. 若在一次试验中，每个基本事件发生的可能性都相同，则称这些基本事件为等可能基本事件.

上面的问题具有以下两个特点：

(1) 所有的基本事件只有有限个；

(2) 每个基本事件的发生都是等可能的.

我们将满足上述条件的随机试验的概率模型称为古典概型(classical probability model).

如果一次试验的等可能基本事件共有 n 个，那么每一个等可能基本事件发生的概率都是$\frac{1}{n}$. 如果某个事件 A 包含了其中 m 个等可能基本事件，那么事件 A 发生的概率为

$$P(A)=\frac{m}{n}.$$

例 1　一只口袋内装有大小相同的 5 只球，其中 3 只白球，2 只黑球，从中一次摸出两只球.

(1) 共有多少个基本事件？

(2) 摸出的两只球都是白球的概率是多少？

分析　可用枚举法找出所有的等可能基本事件.

解　(1) 分别记白球为 1，2，3 号，黑球为 4，5 号，从中摸出 2 只球，有如下基本事件(摸到 1，2 号球用(1，2)表示)：

$$(1,\ 2),\ (1,\ 3),\ (1,\ 4),\ (1,\ 5),\ (2,\ 3),$$
$$(2,\ 4),\ (2,\ 5),\ (3,\ 4),\ (3,\ 5),\ (4,\ 5),$$

因此,共有 10 个基本事件.

(2) 如图 7-2-1,上述 10 个基本事件发生的可能性相同,且只有 3 个基本事件是摸到两只白球(记为事件 A),即(1, 2), (1, 3), (2, 3),故

$$P(A)=\frac{3}{10}.$$

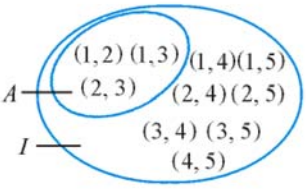


图 7-2-1

答　共有 10 个基本事件,摸出两只球都是白球的概率为$\frac{3}{10}$.

例 2　豌豆的高矮性状的遗传由其一对基因决定,其中决定高的基因记为 D,决定矮的基因记为 d,则杂交所得第一子代的一对基因为 Dd. 若第二子代的 D, d 基因的遗传是等可能的,求第二子代为高茎的概率(只要有基因 D 则其就是高茎,只有两个基因全是 d 时,才显现矮茎).

称 D 为显性基因,d 为隐性基因.

分析　由于第二子代的 D, d 基因的遗传是等可能的,可以将各种可能的遗传情形都枚举出来.

解　Dd 与 Dd 的搭配方式有 4 种:DD, Dd, dD, dd,其中只有第四种表现为矮茎,故第二子代为高茎的概率为$\frac{3}{4}=75\%$.

答　第二子代为高茎的概率为 75%.

思　考

你能求出上述第二子代的种子经自花传粉得到的第三子代为高茎的概率吗?

例 3　将一颗骰子先后抛掷 2 次,观察向上的点数,问:

(1) 共有多少种不同的结果?

(2) 两数之和是 3 的倍数的结果有多少种?

(3) 两数之和是 3 的倍数的概率是多少?

解　(1) 将骰子抛掷 1 次,它出现的点数有 1, 2, 3, 4, 5, 6 这 6 种结果.

先后抛掷 2 次骰子,第 1 次骰子向上的点数有 6 种结果,对每一种结果,第 2 次又都有 6 种可能的结果,于是一共有

$$6\times 6=36$$

种不同的结果.

(2) 第 1 次抛掷,向上的点数为 1, 2, 3, 4, 5, 6 这 6 个数中的某一个,第 2 次抛掷时都可以有两种结果,使两次向上的点数和为 3 的倍数(例如,第 1 次向上的点数为 4,则当第 2 次向上的点数为 2 或 5 时,两次的点数之和都为 3 的倍数),于是共有

$$6\times 2=12$$

种不同的结果.

(3) 因为抛掷 2 次得到的 36 种结果是等可能出现的,记“向上的点数之和是 3 的倍数”为事件 A,则事件 A 的结果有 12 种,故所求的概率为

$$P(A)=\frac{12}{36}=\frac{1}{3}.$$

答　先后抛掷 2 次,共有 36 种不同的结果;点数之和是 3 的倍数的结果共有 12 种;点数之和是 3 的倍数的概率为$\frac{1}{3}$.

思　考

图 7－2－2 直观地给出了例 3 第(2)问中的 12 种结果,你能用此图求出向上的点数之和是 4 的倍数的结果有多少种吗?

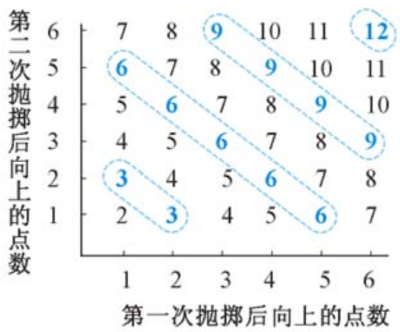


图 7－2－2

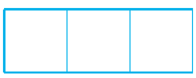


图 7－2－3

例 4　用三种不同颜色给图 7－2－3 中 3 个矩形随机涂色,每个矩形只涂一种颜色,求:

(1) 3 个矩形颜色都相同的概率;

(2) 3 个矩形颜色都不同的概率.

分析　本题中的基本事件较多,为了清楚地枚举出所有可能的基本事件,可画图枚举如下:

这种图叫做树形图,实际上只要画出左边第一个树形图即可推知其余两个的结果.

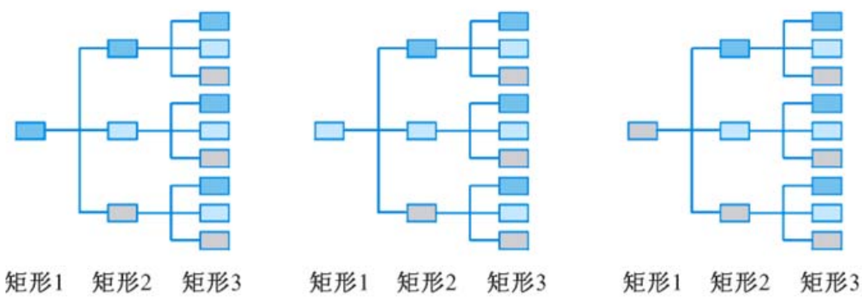


图 7－2－4

解　本题的基本事件共有 27 个(如图 7－2－4).

(1) 记“3 个矩形都涂同一颜色”为事件 A,由图 7－2－4 可知,事件 A 的基本事件有 $1\times3=3$ 个,故

$$P(A)=\frac{3}{27}=\frac{1}{9}.$$

（2）记“3 个矩形颜色都不同”为事件 B，由图 7－2－4 可知，事件 B 的基本事件有 $2\times3=6$ 个，故

$$P(B)=\frac{6}{27}=\frac{2}{9}.$$

答　3 个矩形颜色都相同的概率为$\frac{1}{9}$，3 个矩形颜色都不同的概率为$\frac{2}{9}$.

练　习

1. 某班准备到郊外野营，为此向商店订了帐篷. 如果下雨与不下雨是等可能的，能否准时收到帐篷也是等可能的. 只要帐篷如期运到，他们就不会淋雨，则下列说法中正确的是（　　）.

   A. 一定不会淋雨　　B. 淋雨机会为$\frac{3}{4}$

   C. 淋雨机会为$\frac{1}{2}$　　D. 淋雨机会为$\frac{1}{4}$
2. 一年按 365 天计算，2 名同学在同一天过生日的概率为__________.
3. 一个密码箱的密码由 5 位数字组成，五个数字都可任意设定为 0～9 中的任何一个数字，假设某人已经设定了五位密码.

   （1）若此人忘了密码的所有数字，则他一次就能把锁打开的概率为________；

   （2）若此人只记得密码的前 4 个数字，则一次就能把锁打开的概率为______.
4. 口袋中有形状、大小都相同的 1 只白球和 1 只黑球，先摸出 1 只球，记下颜色后放回口袋，然后再摸出 1 只球.

   （1）一共可能出现多少种不同的结果？

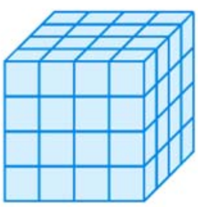
   （2）出现“1 只白球、1 只黑球”的结果有多少种？

   （3）出现“1 只白球、1 只黑球”的概率是多少？
5. 某拍卖行拍卖的 20 幅名画中，有 2 幅是赝品. 某人在这次拍卖中买入了 1 幅画，求买入的这幅画是赝品的概率.

习题 7.2

感受·理解

1. 从含有 500 个个体的总体中，一次性地抽出 25 个个体，假定其中每个个体被抽到的概率相等，那么，总体中某个个体被抽到的概率为______.
2. 某种产品共 100 件，其中有一等品 28 件，二等品 65 件，一等品与二等品都是正品，其余为次品. 某人买了这些产品中的 1 件，问：他买到一等品的概率是多少？买到正品的概率是多少？
3. 如图，把一个体积为 64 cm^3 的正方体木块表面涂上红漆，然后锯成体积为 1 cm^3 的小正方体，从中任取一块，求这一块至少有一面涂有红漆的概率.
4. 连续 3 次抛掷同一颗骰子，求 3 次掷得的点数之和为 16 的概率.



（第 3 题）

5. 100 张卡片上分别写有 1，2，3，…，100，计算：

(1) 任取其中 1 张，这张卡片上写的数是 6 的倍数的结果有多少种？

(2) 任取其中 1 张，这张卡片上写的数是 6 的倍数的概率是多少？

6. 甲、乙、丙三人在 3 天节日中值班，每人值班 1 天，那么甲排在乙前面值班的概率是多少？

7. 某电台访谈节目的听众来自某市的 A,B,C 三个县，主持人从这 3 个县收到的电话数与这三个县的人口数成正比. 估计 A,B,C 三个县的人口数分别为 185 万，81 万和 36 万，试求：

(1) 随机接听一个电话，其来自 A 县的概率；

(2) 这天的第一个电话来自 B 县的概率；

(3) 这天的第一个电话不是来自 C 县的概率.

8. 有红、黄、蓝三种颜色的小旗各 3 面，任取其中 3 面挂于一根旗杆上，求：

(1) 三面旗子全是红色的概率；

(2) 恰有两面旗子是红色的概率.

9. 从一副 52 张的扑克牌(不含大、小王)中抽出一张，求：

(1) 抽出一张 7 的概率；

(2) 抽出一张方块的概率；

(3) 抽出一张方块 7 的概率.

10. 在第 1，2，4，6 路公共汽车都要停靠的一个站(假定没有两辆汽车同时到站)，有一乘客等候第 1 路或第 4 路汽车. 假定各路汽车首先到站的可能性相等，求首先到站的车就是这位乘客所要乘的汽车的概率.

思考・运用

11. 某厂生产的 10 件产品中，有 8 件正品，2 件次品，正品与次品在外观上没有区别. 从这 10 件产品中任意抽检 2 件，计算：

(1) 2 件都是正品的概率；

(2) 1 件是正品，1 件是次品的概率；

(3) 如果抽检的 2 件产品都是次品，则这一批产品将被退货，求这批产品被退货的概率.

12. 一只口袋装有形状、大小都相同的 6 只小球，其中有 2 只白球、2 只红球和 2 只黄球. 从中一次随机摸出 2 只球，试求：

(1) 2 只球都是红球的概率；

(2) 2 只球同色的概率；

(3) “恰有 1 只球是白球的概率”是“2 只球都是白球的概率”的多少倍？

探究・拓展

13. 齐王与田忌赛马，田忌的上马优于齐王的中马，劣于齐王的上马，田忌的中马优于齐王的下马，劣于齐王的中马，田忌的下马劣于齐王的下马. 现各出上、中、下三匹马分组分别进行一场比赛，胜两场以上即为获胜. 如双方均不知对方马的出场顺序，探求田忌获胜的概率.

阅　读

小概率事件

公元1053年,北宋大将狄青奉令征讨南方侬智高叛乱,他在誓师时,当着全体将士的面拿出100枚铜钱说:“我把这100枚铜钱抛向空中,如果钱落地后100枚铜钱都会正面朝上,那么,这次出征定能获胜.”当狄青把100枚铜钱都当众抛出后,竟然全部都是正面向上.狄青又命军士取来100枚铁钉,把这100枚铜钱钉在地上,派兵把守,任人观看.于是宋朝部队军心大振,个个奋勇争先,而侬智高部也风闻此事,军心涣散.狄青终于顺利地平定了侬智高的叛乱.

一枚铜钱抛出落下后,正面与反面向上的概率都是$\frac{1}{2}$,抛2枚铜钱,出现4种等可能基本事件,即

(正,正),(正,反),(反,正),(反,反),

所以2枚铜钱抛出落下后2枚都是正面向上的概率等于$\frac{1}{4}$.再看抛3枚铜钱:这只要在2枚铜钱的四种结果后面增加第三枚铜钱的结果,即

(正,正,正),(正,反,正),(反,正,正),(反,反,正),
(正,正,反),(正,反,反),(反,正,反),(反,反,反),

于是抛3枚铜钱共出现2^3种等可能的基本事件,其中都是正面向上的基本事件只有1个,故3枚铜钱都是正面向上的概率为$\frac{1}{2^3}$.

依此类推,可知抛100枚铜钱,正面全部向上的概率为$\frac{1}{2^{100}}$,这样一个概率非常小的事件竟然发生了,大家觉得很神奇.当然只有归之于神灵的保佑了.

在狄青胜利班师时,命人拔下铁钉,拿起铜钱,发现这100枚铜钱两面都是正面图案,原来这些铜钱是狄青专门铸造的.所以这“100枚铜钱全部正面朝上”的事件是一个必然事件,而在不明内情的人看来,却几乎是不可能事件.狄青正是利用了人们的这个常规想法,使大家都认为几乎不可能发生的事件变成了现实,从而起到了鼓舞本军斗志,涣散敌人士气的巨大作用.

一个事件发生的概率相对很小,就称为小概率事件.小概率事件一般认为在1次试验中几乎不会发生.这就是小概率原理.军事家们利用小概率事件演绎了许多惊心动魄的战例,小说家们利用小概率事件构思了许多引人入胜的情节.日常生活中也有很多小概率事件的例子,你能再举一些吗?

7.3 几何概型

我们来看下面的问题：

(1) 取一根长度为 3 m 的绳子，拉直后在任意位置剪断，那么剪得两段的长都不小于 1 m 的概率有多大?

(2) 射箭比赛的箭靶涂有五个彩色得分环. 从外向内为白色、黑色、蓝色、红色，靶心是金色. 金色靶心叫“黄心”. 奥运会的比赛靶面直径为 122 cm，靶心直径为 12.2 cm. 运动员在 70 m 外射箭. 假设射箭都能中靶，且射中靶面内任一点都是等可能的，那么射中黄心的概率为多少?

在第一个试验中，从每一个位置剪断都是一个基本事件，剪断位置可以是长度为 3 m 的绳子上的任意一点.

在第二个试验中，射中靶面上每一点都是一个基本事件，这一点可以是靶面直径为 122 cm 的大圆内的任意一点.

● 在这两个问题中，基本事件有无限多个，虽然类似于古典概型的“等可能性”还存在着，但是显然不能用古典概型的方法求解. 怎么办呢?

考虑第一个问题，如图 7-3-1，记“剪得两段绳长都不小于 1 m”为事件 A. 把绳子三等分，于是当剪断位置处在中间一段上时，事件 A 发生. 由于中间一段的长度等于绳长的$\frac{1}{3}$，于是事件 A 发生的概率 $P(A)=\frac{1}{3}$.

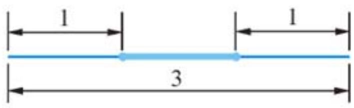


图 7-3-1

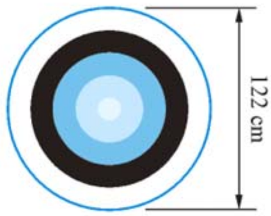


图 7-3-2

再看第二个问题. 如图 7-3-2，记“射中黄心”为事件 B，由于中靶点随机地落在面积为 $\frac{1}{4}\times\pi\times122^2\ \mathrm{cm}^2$ 的大圆内，而当中靶点落在面积为 $\frac{1}{4}\times\pi\times12.2^2\ \mathrm{cm}^2$ 的黄心内时，事件 B 发生，于是事件 B 发生的概率

$$P(B)=\frac{\frac{1}{4}\times\pi\times 12.2^2}{\frac{1}{4}\times\pi\times 122^2}=0.01.$$

从上面的分析可以看到，对于一个随机试验，我们将每个基本事件理解为从某个特定的几何区域内随机地取一点，该区域中每一点被取到的机会都一样；而一个随机事件的发生则理解为恰好取到上述区域内的某个指定区域中的点。这里的区域可以是线段、平面图形、立体图形等。用这种方法处理随机试验，称为几何概型(geometric probability model)。

一般地，在几何区域 D 中随机地取一点，记事件“该点落在其内部一个区域 d 内”为事件 A，则事件 A 发生的概率

$$P(A)=\frac{d\text{ 的测度}}{D\text{ 的测度}}.$$

这里要求 D 的测度不为 0，其中“测度”的意义依 D 确定，当 D 分别是线段、平面图形和立体图形时，相应的“测度”分别是长度、面积和体积等。

例 1　取一个边长为 $2a$ 的正方形及其内切圆(如图 7－3－3)，随机向正方形内丢一粒豆子，求豆子落入圆内的概率。

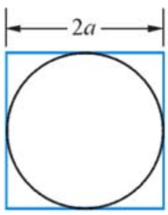


图 7－3－3

分析　由于是随机丢豆子，故可认为豆子落入正方形内任一点都是机会均等的，于是豆子落入圆中的概率应等于圆面积与正方形面积的比。

解　记“豆子落入圆内”为事件 A，则

$$P(A)=\frac{\text{圆面积}}{\text{正方形面积}}=\frac{\pi a^2}{4a^2}=\frac{\pi}{4}.$$

答　豆子落入圆内的概率为 $\frac{\pi}{4}$。

EXCEL

由上例可知，豆子落入圆内的概率 $P(A)=\frac{\pi}{4}$。如果我们向正方形内撒 n 颗豆子，其中落在圆内的豆子数为 m，那么当 n 很大时，比值 $\frac{m}{n}$，即频率应接近于 $P(A)$，于是有

$$P(A)\approx\frac{m}{n}.$$

由此得到

$$\pi\approx\frac{4m}{n}.$$

我们可用 Excel 来模拟撒豆子的试验，以此来估计圆周率，如图 7－3－4 所示。

在 http://www.1088.com.cn/move/003/下载课件.

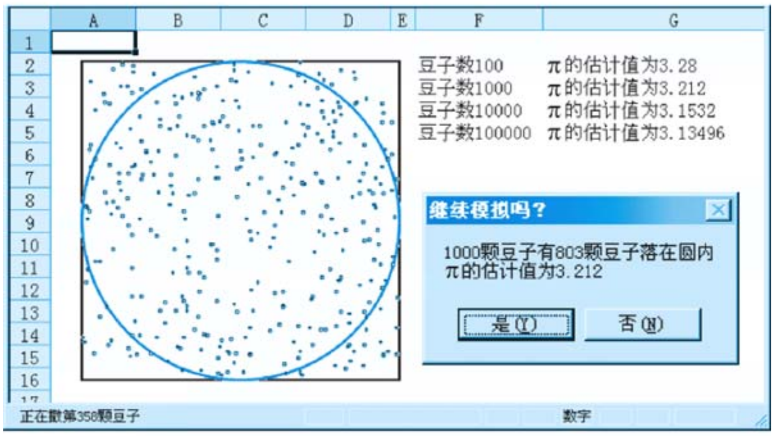


图 7-3-4

例 2 在 1 L 高产小麦种子中混入了一粒带麦锈病的种子，从中随机取出 10 mL，含有麦锈病种子的概率是多少？

分析 病种子在这 1 L 种子中的分布可以看做是随机的，取得的 10 mL种子可视作区域 d，所有种子可视为区域 D.

解 取出 10 mL 麦种，其中“含有病种子”这一事件记为 A，则

$$P(A)=\frac{\text{取出种子的体积}}{\text{所有种子的体积}}=\frac{10}{1\,000}=\frac{1}{100}.$$

答 含有麦锈病种子的概率为$\frac{1}{100}$.

例 3 在等腰直角三角形 ABC 中，在斜边 AB 上任取一点 M，求 AM 小于 AC 的概率.

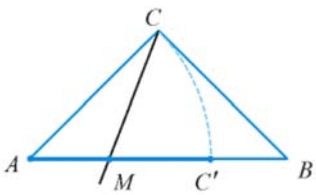


图 7-3-5

分析 点 M 随机地落在线段 AB 上，故线段 AB 为区域 D. 当点 M 位于图 7-3-5 中的线段 AC' 上时，$AM<AC$，故线段 AC' 即为区域 d.

解 在 AB 上截取 $AC'=AC$. 于是

$$\begin{aligned}P(AM<AC)&=P(AM<AC')\\&=\frac{AC'}{AB}=\frac{AC}{AB}=\frac{\sqrt{2}}{2}.\end{aligned}$$

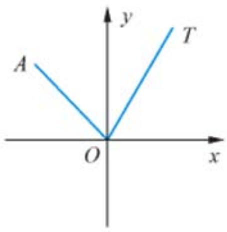
答 AM 小于 AC 的概率为$\frac{\sqrt{2}}{2}$.

练 习

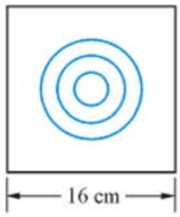
1. 某人睡午觉醒来，发觉表停了，他打开收音机想听电台整点报时，求他等待的时间短于 10 min 的概率.
2. 已知地铁列车每 10 min 一班，在车站停 1 min. 求乘客到达站台立即乘上车的概率.

3. 在 1 万平方公里的海域中有 40 平方公里的大陆架贮藏着石油. 假如在海域中任意一点钻探，钻到油层面的概率是多少？

4. 如图，在直角坐标系内，射线 OT 落在 $60°$ 的终边上，任作一条射线 OA，求射线 OA 落在 $\angle xOT$ 内的概率.



（第 4 题）



（第 5 题）

5. 如图，在墙上挂着一块边长为 16 cm 的正方形木板，上面画了小、中、大三个同心圆，半径分别为 2 cm，4 cm，6 cm，某人站在 3 m 之外向此板投镖. 设投镖击中线上或没有投中木板时都不算，可重投，问：

(1) 投中大圆内的概率是多少？

(2) 投中小圆与中圆形成的圆环的概率是多少？

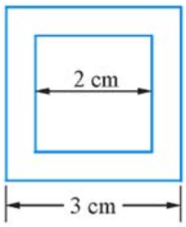
(3) 投中大圆之外的概率是多少？

习题 7.3

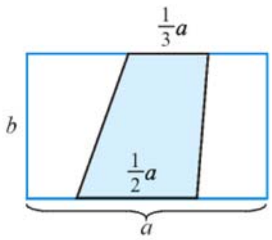
感受 · 理解

1. 两根相距 6 m 的木杆上系一根绳子，并在绳子上挂一盏灯，求灯与两端距离都大于 2 m 的概率.

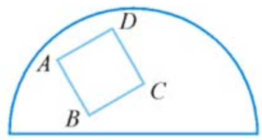
2. 如图，在一个边长为 3 cm 的正方形内部画一个边长为 2 cm 的正方形，向大正方形内随机投点，求所投的点落入小正方形内的概率.



（第 2 题）



（第 3 题）



（第 4 题）

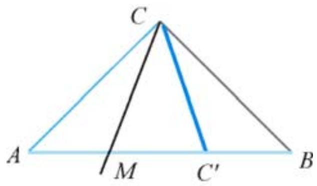
3. 如图，在一个边长为 a，b $(a>b>0)$ 的矩形内画一个梯形，梯形上、下底分别为 $\frac{1}{3}a$ 与 $\frac{1}{2}a$，高为 b. 向该矩形内随机投一点，求所投的点落在梯形内部的概率.

4. 如图，在半径为 1 的半圆内，放置一个边长为 $\frac{1}{2}$ 的正方形 $ABCD$，向半圆内任投一点，求该点落在正方形内的概率.

思考 · 运用

5. 设有一个正方形网格，其中每个最小正方形的边长都等于 6 cm. 现用直径等于 2 cm 的硬币投掷到此网格上，求硬币落下后与格线有公共点的概率.

探究·拓展



（第 6 题）

6. （阅读题）将本节例 3 改为：如图，在等腰直角三角形 ABC 中，过直角顶点 C 在 $\angle ACB$ 内部任作一条射线 CM，与线段 AB 交于点 M，求 $AM<AC$ 的概率. 这时应视作射线 CM 在 $\angle ACB$ 内是等可能分布的，在 AB 上取 $AC'=AC$，则 $\angle ACC'=67.5°$，故满足条件的概率为 $\dfrac{67.5}{90}=\dfrac{3}{4}$.

由此可见，背景相似的问题，当等可能的角度不同时，其概率是不一样的. 请从不同角度研究著名的“贝特朗(Bertrand)问题”：在半径为 1 的圆内随机地取一条弦，则其长超过该圆内接等边三角形的边长的概率是多少？

阅　读

1777 年法国科学家布丰(G. L. L. Buffon，1707～1788)做了一个投针试验，这个试验被认为是几何概型的第一个试验. 他在一张大纸上画了一些平行线，相邻两条平行线间的距离都相等. 再把长度等于平行线间距离一半的针投到纸上，并记录投针的总次数及针落到纸上后与平行线中的某一根相交的次数，共计投针 2 212 次，其中与平行线相交的有 704 次，发现它们的商 $2\,212 \div 704 \approx 3.142\,045$，与 π 非常接近.

以后又有多位数学家重复做过投针试验，得到了类似的结果.

那么，投针试验为什么能算出 π 的近似值呢？

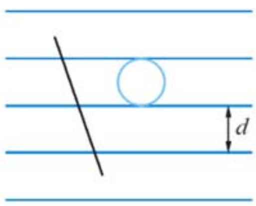


图 7－3－6

如图 7－3－6，取一张大纸，在上面画一组平行线，使相邻两平行线距离都等于 d，再取一个直径为 d 的铁丝圆圈. 如果把这个铁丝圆圈投掷到纸上，则圆圈与平行线组的交点肯定是 2 个. 如果投掷 n 次，则交点总计应为 $2n$. 如果把铁丝拉直(长度不变)再投掷到纸上，则铁丝与平行线组的交点就可能是 4 个、3 个、2 个、1 个或 0 个.

布丰认为，既然两根铁丝长度相等，在大量重复试验时，它们与同一组平行线的交点总数应是相等的. 如果也投掷 n 次，则交点总计也应与 $2n$ 相差甚小. 再考虑铁丝上的每个点，它是否落在平行线组的某一根上也是机会均等的.

现在如果取一段长为 l 的铁丝，则投掷 n 次时，交点总数 n' 应与 $\dfrac{l}{\pi d}\cdot 2n$ 相差甚小. 即如果一根长度为 l 的铁丝投掷 n 次，得到交点总数为 $n'\approx\dfrac{l}{\pi d}\cdot 2n$. 故当投掷次数 n 较大时，数 $\dfrac{2nl}{n'd}$ 应在 π 附近摆动，布丰取 $l=\dfrac{1}{2}d$，则 $\dfrac{n}{n'}$ 应在 π 附近摆动. 布丰试验的结果正好反映了这一事实.

7.4 互斥事件及其发生的概率

● 体育考试的成绩分为四个等级：优、良、中、不及格，某班 50 名学生参加了体育考试，结果如下：

表 7-4-1

优	85 分及以上	9 人
良	75～84 分	15 人
中	60～74 分	21 人
不及格	60 分以下	5 人

体育考试的成绩的等级为优、良、中、不及格的事件分别记为 A，B，C，D.

(1) 在同一次考试中，某一位同学能否既得优又得良？

(2) 从这个班任意抽取一位同学，那么这位同学的体育成绩为"优良"(优或良)的概率是多少？

在同一次体育考试中，同一人不可能既得优又得良，即事件 A 与 B 是不可能同时发生的. 不能同时发生的两个事件称为互斥事件(exclusive events).

对于本例中的事件 A，B，C，D，其中任意两个都是互斥的. 一般地，如果事件 A_1，A_2，…，A_n 中的任何两个都是互斥事件，就说事件 A_1，A_2，…，A_n 彼此互斥.

设 A，B 为互斥事件，当事件 A，B 有一个发生，我们把这个事件记作 $A+B$. 在上述关于体育考试成绩的问题中，事件 $A+B$ 就表示事件"优"或"良"，那么，事件 $A+B$ 发生的概率是多少呢？

从 50 人中任意抽取 1 个人，有 50 种等可能的方法，而抽到优良的同学的方法有 $9+15$ 种，从而事件 $A+B$ 发生的概率

$$P(A+B)=\frac{9+15}{50}.$$

另一方面

$$P(A)=\frac{9}{50},\ P(B)=\frac{15}{50},$$

因此有

$$P(A+B)=P(A)+P(B).$$

由以上分析不难发现，概率必须满足如下第三个基本要求：

如果事件 A，B 互斥，那么事件 $A+B$ 发生的概率，等于事件 A，B 分别发生的概率的和，即

$$P(A+B)=P(A)+P(B).$$

一般地，如果事件 A_1，A_2，…，A_n 两两互斥，则

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n).$$

在上面的问题中，如果将“体育成绩及格”记为事件 E，那么 E 与 D 不可能同时发生，但必有一个发生.

两个互斥事件必有一个发生，则称这两个事件为对立事件(complementary events). 事件 A 的对立事件记为 $\overline{A}$.

思　考

对立事件与互斥事件有何异同？

对立事件 A 与 $\overline{A}$ 必有一个发生，故 $A+\overline{A}$ 是必然事件，从而

$$P(A)+P(\overline{A})=P(A+\overline{A})=1.$$

由此，我们可以得到一个重要公式：

$$P(\overline{A})=1-P(A).$$

例 1　一只口袋内装有大小一样的 4 只白球与 4 只黑球，从中一次任意摸出 2 只球. 记摸出 2 只白球为事件 A，摸出 1 只白球和 1 只黑球为事件 B. 问：事件 A 与 B 是否为互斥事件？是否为对立事件？

解　事件 A 与 B 互斥.

因为从中一次可以摸出 2 只黑球，所以事件 A 与 B 不是对立事件.

例 2　某人射击 1 次，命中 7～10 环的概率如表 7-4-2 所示：

表 7-4-2

命中环数	10　环	9　环	8　环	7　环
概　　率	0.12	0.18	0.28	0.32

(1) 求射击 1 次，至少命中 7 环的概率；

(2) 求射击 1 次，命中不足 7 环的概率.

解　记事件“射击 1 次，命中 k 环”为 A_k($k\in\mathbf{N}$，且 $k\leqslant 10$)，则事件 A_k 彼此互斥.

(1) 记“射击一次，至少命中 7 环”的事件为 A，那么当 A_{10}，A_9，

A_8 或 A_7 之一发生时，事件 A 发生. 由互斥事件的概率加法公式，得

$$\begin{aligned}P(A) &= P(A_{10}+A_9+A_8+A_7)\\ &= P(A_{10})+P(A_9)+P(A_8)+P(A_7)\\ &= 0.12+0.18+0.28+0.32\\ &= 0.9.\end{aligned}$$

(2) 事件“射击一次，命中不足 7 环”是事件“射击一次，命中至少 7 环”的对立事件，即$\overline{A}$表示事件“射击一次，命中不足 7 环”. 根据对立事件的概率公式，得

$$P(\overline{A}) = 1-P(A) = 1-0.9 = 0.1.$$

答　此人射击 1 次，至少命中 7 环的概率为 0.9；命中不足 7 环的概率为 0.1.

例 3　黄种人群中各种血型的人所占的比如表 7 - 4 - 3 所示：

表 7 - 4 - 3

血　　型	A	B	AB	O
该血型的人所占比/%	28	29	8	35

已知同种血型的人可以输血，O 型血可以输给任一种血型的人，任何人的血都可以输给 AB 型血的人，其他不同血型的人不能互相输血. 小明是 B 型血，若小明因病需要输血，问：

(1) 任找一个人，其血可以输给小明的概率是多少？

(2) 任找一个人，其血不能输给小明的概率是多少？

解　(1) 对任一人，其血型为 A，B，AB，O 型血的事件分别记为 A'，B'，C'，D'，它们是互斥的. 由已知，有

$$P(A') = 0.28,\ P(B') = 0.29,$$

$$P(C') = 0.08,\ P(D') = 0.35.$$

因为 B，O 型血可以输给 B 型血的人，故“可以输给 B 型血的人”为事件 $B'+D'$. 根据互斥事件的加法公式，有

$$\begin{aligned}P(B'+D') &= P(B')+P(D')\\ &= 0.29+0.35 = 0.64.\end{aligned}$$

(2) 由于 A，AB 型血不能输给 B 型血的人，故“不能输给 B 型血的人”为事件 $A'+C'$，且

$$\begin{aligned}P(A'+C') &= P(A')+P(C')\\ &= 0.28+0.08 = 0.36.\end{aligned}$$

答 任找一人,其血可以输给小明的概率为 0.64,其血不能输给小明的概率为 0.36.

第(2)问也可以这样解:因为事件“其血可以输给 B 型血的人”与事件“其血不能输给 B 型血的人”是对立事件,故由对立事件的概率公式,有

$$\begin{aligned} P(\overline{B'+D'}) &= 1-P(B'+D') \\ &= 1-0.64=0.36. \end{aligned}$$

练　习

1. 抛掷一颗骰子 1 次,记“向上的点数是 4,5,6”为事件 A,“向上的点数是 1,2”为事件 B,“向上的点数是 1,2,3”为事件 C,“向上的点数是 1,2,3,4”为事件 D. 判断下列每对事件是否为互斥事件,如果是,再判断它们是否为对立事件.

(1) A 与 B;　　(2) A 与 C;　　(3) A 与 D.

2. 从装有 5 只红球、5 只白球的袋中任意取出 3 只球,有事件:① “取出 2 只红球和 1 只白球”与“取出 1 只红球和 2 只白球”;② “取出 2 只红球和 1 只白球”与“取出 3 只红球”;③ “取出 3 只红球”与“取出 3 只球中至少有 1 只白球”;④ “取出 3 只红球”与“取出 3 只白球”. 其中是对立事件的有(　　).

A. ①、④　　B. ②、③　　C. ③、④　　D. ③

3. 有一批小包装食品,其中重量在 90～95 g 的有 40 袋,重量在 95～100 g 的有 30 袋,重量在 100～105 g 的有 10 袋. 从中任意抽取 1 袋,则此袋食品的重量在 95～100 g 的概率为______;此袋食品的重量不足 100 g 的概率为______;此袋食品的重量不低于 95 g 的概率为______.(重量在 a～b g 指的是重量的数值在区间$[a, b)$内)

4. 某地区年降水量(单位:mm)在下列范围内的概率如下表:

年降水量	[600, 800)	[800, 1 000)	[1 000, 1 200)	[1 200, 1 400)	[1 400, 1 600)
概　率	0.12	0.26	0.38	0.16	0.08

(1) 求年降水量在[800, 1 200)内的概率;

(2) 如果年降水量≥1 200 mm,就可能发生涝灾,求该地区可能发生涝灾的概率.

习题 7.4

感受 · 理解

1. 口袋中有若干红球、黄球与蓝球,摸出红球的概率为 0.45,摸出黄球的概率为 0.33,求:

(1) 摸出红球或黄球的概率;

(2) 摸出蓝球的概率.

2. 一架飞机向目标投弹,击毁目标的概率为 0.2,目标未受损的概率为 0.4,求使目标受损但未完全击毁的概率.

3. 经统计,在某储蓄所一个营业窗口等候的人数及相应概率如下:

排队人数	0	1	2	3	4	5人及5人以上
概　率	0.1	0.16	0.3	0.3	0.1	0.04

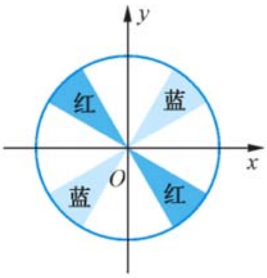
(1) 至多2人排队等候的概率是多少?

(2) 至少3人排队等候的概率是多少?

4. 某种彩色电视机的一等品率为90%,二等品率为8%,次品率为2%,某人买了一台该种彩色电视机,求:

(1) 这台电视机是正品(一等品或二等品)的概率;

(2) 这台电视机不是一等品的概率.



(第5题)

5. 在直角坐标系中画一个直径为40 cm的圆,以各象限的角平分线为对称轴画四个30°的扇形,并涂以红、蓝两色,其余部分涂以白色(如图).现用一支小镖投向圆面,假定都能投中圆面,求:

(1) 分别投中红色、蓝色扇形区域的概率;

(2) 投中红色或蓝色扇形区域的概率;

(3) 投中白色扇形区域的概率.

6. 经临床验证,一种新药对某种疾病的治愈率为54%,显效率为22%,有效率为12%,其余为无效.求某人患该病使用此药后无效的概率.

思考·运用

7. 将扑克牌四种花色的A,K,Q共12张,洗匀.

(1) 甲从中任意抽取2张,求抽出的2张都为A的概率;

(2) 若甲已抽到了2张K后未放回,求乙抽到2张A的概率.

探究·拓展

8. 某种彩票是由7位数字组成,每位数字均为0～9这10个数码中的任一个.由摇号得出一个7位数(首位可为0)为中奖号,如果某张彩票的7位数与中奖号相同即得一等奖;若有6位相连数字与中奖号的相应数位上的数字相同即得二等奖;若有5位相连数字与中奖号的相应数位上的数字相同即得三等奖;各奖不可兼得.某人一次买了10张不同号码的彩票.

(1) 求其获得一等奖的概率;

(2) 求其获得三等奖及以上奖的概率.

阅　读

制作一个如图7-4-1所示的通道及下方相互隔离的储槽.

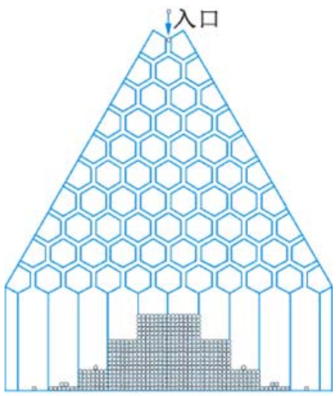


图7-4-1

若以 1 颗球形小珠放入最上方的通道入口，那么小珠落入下方每个储槽的概率有何规律？

在上方的入口箭头所指处投入 2^n 粒球形小珠，则每粒小珠落下时，每到一层的分岔处，都以 $\frac{1}{2}$ 的概率向左边或右边的下一层垂直通道下落.

显然，任何一层的最左、最右的两个通道都只有 1 个可能情况，而每一层的其他通道通过小珠的可能情况，应等于其两肩上两个通道的可能情况的和. 这样就得到了图 7－4－2 所示的数表：

$$
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \\
1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1 \\
1 \quad 9 \quad 36 \quad 84 \quad 126 \quad 126 \quad 84 \quad 36 \quad 9 \quad 1
\end{array}
$$

图 7－4－2

由此，当通道共 9 层时，放入 2^9 粒小珠后在从左至右的 10 个储槽中落入小珠的数目，按可能情况计算，应是

$$1，9，36，84，126，126，84，36，9，1.$$

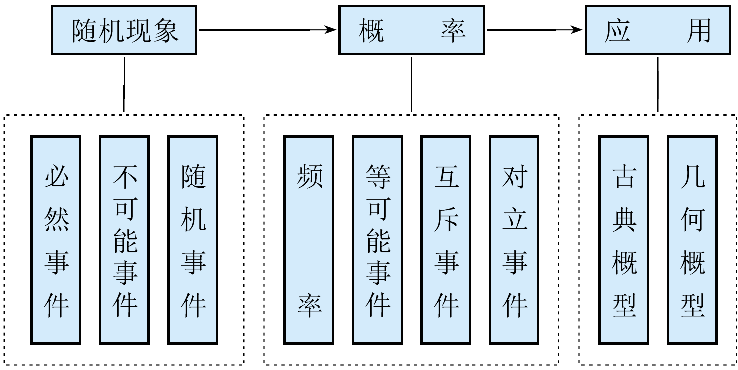
故所求概率依次为

$$\frac{1}{2^9},\ \frac{9}{2^9},\ \frac{36}{2^9},\ \frac{84}{2^9},\ \frac{126}{2^9},\ \frac{126}{2^9},\ \frac{84}{2^9},\ \frac{36}{2^9},\ \frac{9}{2^9},\ \frac{1}{2^9}.$$

图 7－4－2 所示的数表最早由我国宋朝时期的杨辉于 1261 年画在他所著的《详解九章算法》中，并说明出于《释锁算书》，贾宪曾用过此图，西方称此图为帕斯卡三角形，是帕斯卡于 1654 年发表的，但更早的德国人阿批纳斯(Pertrus Apianus)于 1527 年便把这张图刻在他的一本算术书上了.

本章回顾

我们通过对随机现象的分析，了解了概率的含义，运用等可能事件、互斥事件与对立事件的性质，研究了古典概型和几何概型及一些简单概率的计算方法.



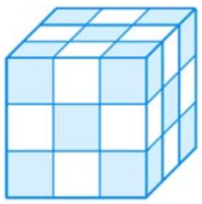
对随机事件进行大量重复试验时，其发生的频率稳定在某个常数附近，这个常数就反映了随机事件发生的可能性的大小，我们用“概率”加以描述.

运用等可能事件、互斥事件及对立事件的性质，可以根据随机事件的内在规律确定其发生的概率，这为解决古典概型、几何概型等概率问题提供了理论依据. 我们要善于将实际问题转化为古典概型或几何概型.

复习题

感受·理解

1. 某班要选一名学生做代表，每个学生当选是等可能的，若“选出代表是男生”的概率是“选出代表是女生”的概率的$\frac{4}{5}$，求这个班的男生人数占全班人数的百分比.



（第2题）

2. 如图，边长为1的红色小正方体与白色小正方体相间堆成一个$3\times3\times3$大正方体（同色正方体都没有相邻的面），从中任选一个小正方体，选中红色正方体的概率是多少？

3. 在10件产品中，有一等品7件，二等品2件（一等品与二等品都是正品），次品1件，现从中任取2件.

(1) 两件都是一等品的概率是多少？两件都是二等品的概率是多少？

(2) 两件中有1件是次品的概率是多少？

(3) 两件都是正品的概率是多少？

4. 某单位要在4名工人中安排2名分别到两处出差（每人被安排是等可能的）.

(1) 共有多少种安排方法？

(2) 其中甲、乙两人都被安排的方法有多少种？

(3) 其中甲、乙两人都被安排的概率是多少？

5. 抛掷一枚硬币3次，分别求掷得0次、1次、2次、3次正面向上的概率.

6. 连续抛掷一颗骰子2次，分别求掷出的点数和为2，3，…，12时的概率.

7. 有五条线段，其长度分别为1，3，5，7，9. 现任取三条，求能构成三角形的概率.

8. 用计算机随机产生的有序二元数组(x, y)，满足$-1<x<1$，$-1<y<1$，对每个有序二元数组(x, y)，用计算机计算x^2+y^2的值，记A为事件“$x^2+y^2<1$”. 试求事件A发生的概率.

思考·运用

9. 一次口试，每位考生要在8道口试题中随机抽出2道题回答，若答对其中1题即为及格.

(1) 现有某位考生会答8道题中的5道题，那么，这位考生及格的概率有多大？

(2) 如果一位考生及格的概率小于50%，则他最多只会几道题？

10. 国家安全机关用监听录音机记录了两个间谍的谈话，发现30 min长的磁带上，从开始30 s处起，有10 s长的一段内容包含两间谍犯罪的信息. 后来发现，这段谈话的一部分被某工作人员擦掉了，该工作人员声称她完全是无意中按错了键，使从此处起往后的所有内容都被擦掉了. 那么由于按错了键使含有犯罪内容的谈话被部分或全部擦掉的概率有多大？

探究·拓展

11. 两个水平相当的选手在决赛中相遇，决赛采用五局三胜制，胜者获得全部奖金，前3局打成2∶1时比赛因故终止. 有人提出按2∶1分配奖金，你认为这样分配合理吗？为什么？

附录 1：随机数表（部分）

03 47 43 73 86	36 96 47 36 61	46 98 63 71 62	33 26 16 80 45	60 11 14 10 95
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51	24 51 79 89 73
16 76 62 27 66	56 50 26 71 07	32 90 79 78 53	13 55 38 58 59	88 97 54 14 10
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21	88 26 49 81 76
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53	23 83 01 30 30
16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79
33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28
18 18 07 92 45	44 17 16 58 09	79 83 86 19 62	06 76 50 03 10	55 23 64 05 05
26 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45	10 93 72 88 71
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72	93 85 79 10 75
52 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53	86 60 42 04 53
37 85 94 35 12	83 39 50 08 30	42 34 07 96 88	54 42 06 87 98	35 85 29 48 39
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04	07 74 21 19 30
56 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54	97 77 46 44 80
99 49 57 22 77	88 42 95 45 72	16 64 36 16 00	04 43 18 66 79	94 77 24 21 90
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45	99 27 72 95 14
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 66 60 44	38 68 88 11 80
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52	68 07 97 06 57
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 76	15 54 55 95 52
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45	97 60 49 04 91
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26	11 04 96 67 24
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27	40 48 73 51 92
16 90 82 66 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68	02 02 37 03 31
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86	38 45 94 30 38
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16	02 75 50 95 98
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62	48 51 84 08 32
31 96 25 91 47	96 44 33 49 13	34 86 82 53 91	00 52 43 48 85	27 55 26 89 62
66 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90	57 16 00 11 66
14 90 84 45 11	76 73 88 05 90	52 27 41 14 86	22 98 12 22 08	07 52 74 95 80
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90	49 37 38 44 59
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36	47 95 93 13 30
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 78 89 62	02 67 74 17 33
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83	52 91 05 70 74
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93	58 05 77 09 51
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16	29 56 24 29 48
26 99 61 65 53	58 37 78 80 70	42 10 50 67 42	32 17 55 85 74	94 44 67 16 94
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50	15 29 39 39 43

计算器使用范例

范例 1：小数位数/有效位数

(1)　200 ÷ 7 × 14 = 400

200 ÷ 7 × 14　　=　400

（指定3位小数）MODE MODE MODE 1 3

400.000

(2)　2 ÷ 7，以 3 位有效位数（SCI3）显示计算结果．

MODE MODE MODE 2 3

2 ÷ 7　=　2.86^{-01}

注：若要恢复请按 MODE MODE MODE 3 1

范例 2：利用变数进行计算

已知方程　$3.2x^2-9.2x+4.7=0$，试根据公式

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

求解方程．

3.2 SHIFT STO A (−) 9.2 STO SHIFT B

4.7 SHIFT STO C AC ALPHA

B X^2 − 4 ALPHA A ALPHA

C SHIFT STO D AC ((−) ALPHA

B + √ ALPHA D) ÷

(2 ALPHA A)

=　2.210582305

按◀键直到B与√之间，即+下方，改为−

=　0.664417695

范例 3：统计计算

考察某学校学生上课迟到的情况，该学校 2 308 个学生半年上课迟到次数列表如下，求总体平均数、方差．

迟到人数	0	1	2	3	4	5
人　　数	557	503	483	375	232	158

解：按 MODE 2　（进入统计状态）

SHIFT CLR 1（Scl）（消除存储器内容）

AC 0 SHIFT ; 557 DT 1 SHIFT ; 503 DT

2 SHIFT ; 483 DT 3 SHIFT ; 375 DT

4 SHIFT ; 232 DT 5 SHIFT ; 158 DT

SHIFT S-VAR 1 （$\bar{x}$） =　$\bar{x}$ 1.868284229

SHIFT S-VAR 2 （$x\sigma n$） =　$x\sigma n$ 1.531862405

X^2 =　Ans^2 2.346602429

说　　明

江苏教育出版社出版的《普通高中课程标准实验教科书·数学》是根据2003年教育部制订的《普通高中数学课程标准(实验)》编写的.

该套教科书充分体现数学课程标准的基本理念,使学生通过高中阶段的学习,能获得适应现代生活和未来发展所必需的数学素养,满足他们个人发展与社会进步的需要.

教科书力图使学生在丰富的、现实的、与他们经验紧密联系的背景中感受数学、建立数学、运用数学,做到"入口浅,寓意深".通过创设恰当的问题情境,引导学生进行操作、观察、探究和运用等活动,感悟并获得数学知识与思想方法.在知识的发生、发展与运用过程中,培养学生的思维能力、创新意识和应用意识.

教科书按知识发展、背景问题、思想方法三条主线,通过问题将全书贯通.每个模块围绕中心教育目标展开,每章围绕核心概念或原理展开.教科书充分关注数学与自然、生活、科技、文化、各门学科的联系,让学生感受到数学与外部世界是息息相通、紧密相连的.

教科书充分考虑学生的不同需求,为所有学生的发展提供帮助,为学生的不同发展提供较大的选择空间.整个教科书设计为:一个核心(基本教学要求),多个层次,多种选择.学好核心内容后,根据需要,学生有多种选择,每一个人都能获得必备的数学素养与最佳发展.

众多的数学家、心理学家、学科教育专家、特级教师参加了本套教科书的编写工作.参与本册讨论与审稿的专家与教师有:葛军,陈光立,仇炳生,樊亚东,董林伟,陆云泉,寇恒清,于明,孙旭东,张松年,陈跃辉,陶维林,潘娉姣,周焕山,洪再吉等,在此向他们深表感谢!

本书编写组

2004年11月